A STUDY OF THE NORMATIVE PSYCHOMETRIC PROPERTIES OF THE ALFIE, A NOVEL MEASURE OF ACCELERATED LONG-TERM FORGETTING IN TEMPORAL LOBE EPILEPSY

EMMA CAMERON, BSc (Hons)

A thesis submitted in partial fulfilment of the requirements of
The University of Lincoln for the degree of Doctor in Clinical Psychology

2015

Thesis abstract

Background: Patients with Temporal Lobe Epilepsy (TLE) typically complain of memory difficulties, but these are not always evident on objective memory assessments. This discrepancy may result from under- or over-reporting of memory problems as a result of motivation (e.g., to avoid confrontation, functional consequences on employment), impact of psychosocial factors (e.g., mood) or epilepsy patients' perception of their difficulties (e.g., memory disturbances may cause difficulties with patients' metamemory, meaning they lack the ability to accurately comment on their cognitive capabilities). Alternatively, this could be explained by a lack of specificity in current memory measures that utilise delay periods of 30-40 minutes. Research suggests that extending published memory tests to assess recall over a longer period (e.g., a two-week delay) allows the memory deficits described by those with TLE to be seen – a phenomenon known as Accelerated Long-term Forgetting (ALF). However, the literature has yet to address difficulties with test ecological validity and clinical practicability.

Aim: The aim of this research was to further develop a novel measure of ALF to examine whether it was a clinically viable test measure. This was explored through assessments of validity, reliability, and acceptability/practicability of the implemented test procedure, for this novel measure.

Method: 50 healthy participants' objective memory performance was assessed by asking them to recall and recognise information at three time points: immediately after presentation of stimuli (T1), after 40 minutes (T2), and after two-weeks (T3). Alongside a published story and word list tasks, a novel measure was used – the Accelerated Long-term Forgetting In Epilepsy (ALFIE) test. We believe this to be more ecologically valid and clinically practical than other memory measures, due to the test using multi-modal stimuli drawn from real-life televised news broadcasts (concordant with a verisimilitude approach), with use of telephone follow-up phone-calls to assess two-week recall and recognition. Subjective memory performance was assessed via use of a self-report questionnaire.

Results: Convergence of results on the ALFIE and published memory measures was found. Although extension of the published measure to a two-week delay might then seem justifiable, the ALFIE test showed greater correlations with subjective memory scores than the published story and word lists tasks. This suggests greater ecological validity of the ALFIE measure than the published memory test. Reliability was assessed through inter-rater reliability and analysis of parallel forms. The ALFIE showed high inter-rater reliability and although parallel forms reliability was poor, through standardisation versions may be used as alternate forms. Low attrition rates suggest that use of a two-week delayed assessment via telephone might be a clinically viable solution for specialist Epilepsy services assessing ALF, often over large geographical regions (where it would be costly, impractical and hard to co-ordinate for patients to return for extended delay follow-ups within such a restrictive time limit). Significant differences in performance between genders needs further examination, but may partially be explained through emotional salience of materials.

Conclusion: The ALFIE test appears to be a viable test measure for assessing memory that is more ecologically-valid and clinically practicable than current memory measures.

Acknowledgements

This thesis would not have been possible without the time and support of many people. I would firstly like to thank all of the participants who took part in the research. I greatly appreciated the time they gave to be involved.

To my research tutors, Dr Nima Moghaddam and Dr Roshan das Nair, I gratefully acknowledge the supervision, advice and support they gave throughout the research project. Extended thanks go to Dr Nima Moghaddam for always being available at times of statistical stress.

Finally, I would like to thank my family and friends for their practical and emotional support. To my parents, who have always supported my endeavours, and to my partner, Matt, for his love, encouragement and patience. It is so appreciated. Thank you.

Statement of contribution

I, Emma Cameron, Trainee Clinical Psychologist, confirm that I designed this systematic literature review, independently decided upon suitable search terminology, implemented all of the literature search process and reviewed all of the search results, including designing an appropriate critical appraisal tool, and wrote this literature review. My colleagues, Dr Nima Moghaddam and Dr Roshan das Nair, contributed to this review by discussing my initial review question with me and reviewing a journal draft.

I also confirm that I designed this research study, sought and obtained the relevant ethical approval, reviewed the literature pertinent to the research area, was responsible for participant recruitment, conducted all of the data collection and data analysis, including scoring and inputting of the data and interpreting statistical findings, and wrote this thesis. Dr Faye Corbett contributed to this research by providing access to the Accelerated Long-term Forgetting In Epilepsy (ALFIE) test materials and existing clinical population data – both from her own doctoral thesis. Dr Nima Moghaddam and Dr Roshan das Nair, contributed to this research by conceptualising and discussing the research design with me, reviewing the study progress periodically, discussing the research findings and analysis, and reviewing a journal draft.

Contents

Extended Paper

Poster

Small Scale Research & Impact

Tables and figures

Appendices

Systematic Review

The impact of the psychometric assessment process on the client, therapist and therapeutic process: A critical review.

Emma Cameron[a]

Nima Moghaddam[ab]

Roshan das Nair[cd]

[a]Trent Doctorate in Clinical Psychology, School of Life Sciences, Bridge House, Lincoln University, Brayford Campus, Lincoln, LN6 7DL, United Kingdom

[b]Corresponding author: Email: NMoghaddam@post01.lincoln.ac.uk; Telephone: +44 (0) 1522 886 029

[c]Trent Doctorate in Clinical Psychology, Division of Psychiatry and Applied Psychology, School of Medicine, YANG Fujia Building, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, United Kingdom

[d]Dept of Clinical Psychology and Neuropsychology, Nottingham University Hospitals NHS Trust, Nottingham NG7 2UH, United Kingdom

Abstract

Introduction: Many mental-health services use psychometric assessments to diagnose clients and evaluate treatment and service effectiveness. This critical review investigated the impact of the psychometric assessment process on the client, therapist and therapeutic process.

Method: Journal articles were systematically searched through electronic databases. Methodological quality of studies was assessed with a critical appraisal tool developed by the first author.

Results: From a potential 756 studies, 16 studies fulfilled the inclusion criteria. Analysis of retrieved studies was limited by a lack of information around sample eligibility criteria and characteristics, and poor rigour and reflexivity within those studies that applied qualitative methods. Notwithstanding these limits, analysis across studies identified potential utility of psychometric assessments in terms of: opening the way for dialogue between client and therapist, increasing client understanding and insight into their difficulties, and increasing professional empathy.

Conclusion: There is a paucity of well-designed studies in this area. The heterogeneity of relevant studies limited comparison, as much data had to be extrapolated.

Key words: Psychometric assessments, Therapy, Therapeutic process, Critical Review.

Introduction

There are several types of psychometric assessments. They vary in terms of completion times; for example: “ultra-brief” scales such as the Session Rating Scale (SRS; Miller & Duncan, 2004), questionnaires such as the Beck Depression Inventory – second edition (BDI-II; Beck, Steer & Brown, 1996); and longer assessments/batteries such as the Wechsler Adult Intelligence Scale–fourth edition (WAIS-IV; Wechsler, 2008) or the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001). Ultra-brief scales typically take less than five minutes to complete, score and analyse, and are often used to measure a presence of change between sessions, rather than give the precision and depth of information that a longer assessment/battery would. Some assessments focus on one dimension of mental state (e.g., BDI-II), whereas others are multi-dimensional (e.g., WAIS-IV). They can be self-administered (e.g. the SRS or the BDI-II), or need an administrator (e.g., the WAIS-IV or the D-KEFS).

National Institute for Health and Clinical Excellence (NICE) guidelines recommend the use of psychometric assessments as routine practice when working with a variety of mental health problems. For instance, in depression, NICE (2009) states that assessment should be supported by a formal rating tool such as the BDI-II (Beck et al., 1996). Such assessments have been important in determining the effectiveness of therapies; through use of pre- and post-measures, session-by-session ratings, or even comparisons between groups.

Limited research has examined the effects of the assessment process, particularly the actual administration of psychometric tests to clients. Stimson (1974) postulated that allowing clients to share their views of their difficulties via assessment could empower them and increase adherence to, and satisfaction with, treatment; however, this study is dated and assessment measures and practises are constantly being revised.

The process of feeding back or sharing test results with clients is slightly more well-researched, and findings from this have generally been positive, indicating for example, an increase in the collaborative nature of therapy (Tharinger et al., 2008) or an increase in therapeutic alliance (Hilsenroth, Peters, & Ackerman, 2004). A limitation of these ‘feedback studies’ is that a large

proportion of them have been conducted on feeding back to parents and/or children, which is a commonplace procedure in children's services. Research of the effects on adults is not as in-depth. As the literature suggests that a good therapeutic alliance and competence of the therapist can both predict good treatment outcome (Diamond et al., 2006; Horvath & Symonds, 1991; Kuyken & Tsivrikos, 2009), the impact of having to administer (for the professional) and complete (for the client) tests seems particularly salient.

This review therefore aimed to provide a critical synthesis of the literature in this area to: (1) determine the perceived effects of the psychometric assessment process (administration and feedback) on the therapist, the client, and the therapeutic process; and (2) to uncover the areas of research in this field that are lacking and need further examination.

Method

Database selection

Studies were selected for review through database searching. Three databases were used for searching; PsycINFO, MEDLINE and EMBASE. These databases were chosen due to the extensive number of journal articles they are able to access, and the large time period that they cover.

Database search

A search was constructed to run simultaneously in EMBASE, MEDLINE and PsycINFO using the OvidSP (2013) search engine (see Table 1). The search focused on the three key concepts of the literature review question: (1) usefulness, (2) psychometrics, and (3) therapy. For each search concept, a range of terms and synonyms was used to ensure breadth of results; variations on terms were captured using truncation symbols (e.g., 'Therap$' to include 'Therapy', 'Therapies' and 'Therapeutic'). Terms mapping on to the same concept were combined using the Boolean operator 'or.'

In order to apply exclusion criteria to the article search, the Boolean operator 'not' was also used. Firstly, the search term 'Psycholog$' was combined with each of the 'not' criteria to exclude studies that: discussed drug or medication treatment, used a child population sample, or pertained to surgical treatment. Secondly, these criteria were combined together with 'and' to ensure that all of the papers that did not meet the eligibility criteria were excluded. Careful consideration was given to the exclusion terms used and potential limiting effects on sensitivity were minimised through subsequent trawling of reference lists in screened journal articles.

Search terms for individual concepts were then combined using 'and' in order to identify articles that addressed all key concepts together (whilst applying exclusion criteria outlined above). Additional limits were then applied, using functions available on the site, to reflect the eligibility criteria, for example ensuring only human participants within the adult age range. The search results were then 'deduplicated' to remove duplicate studies.

Table 1

Search composition and justification of terms

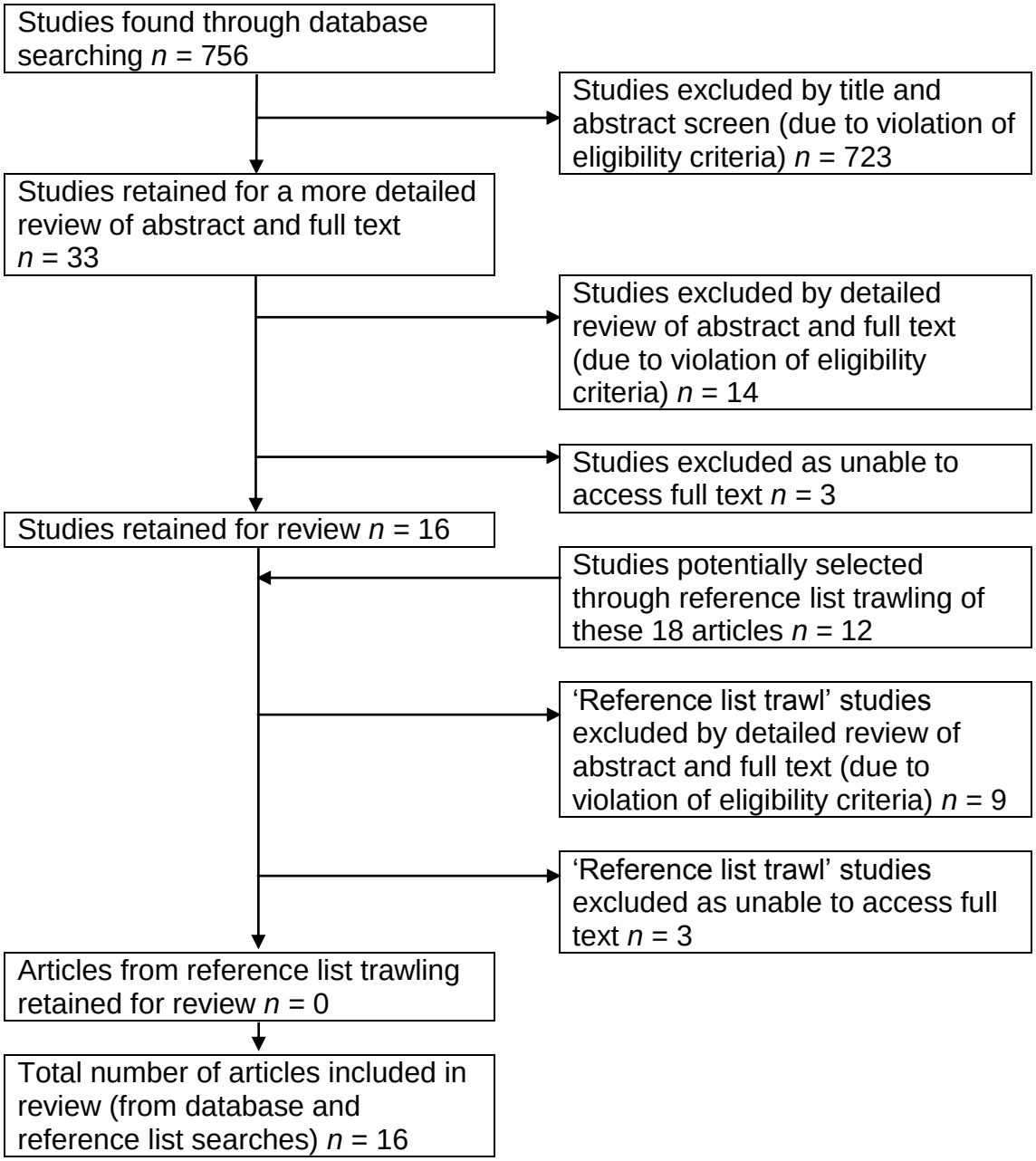
No.	Search Term		Justification		No. of results
			To include:		
1	Utility		Utility		297296
2	Useful$		Useful, Usefulness		1512512
3	Benefi$		Benefit, Benefits, Beneficial		1484926
4	1 or 2 or 3		To include any of the search terms relating to utility		3159835
5	Psychometric$		Psychometric, Psychometrics		156144
6	Psycholog$ assessment$		Psychology assessment, psychological assessment, Psychology assessments, psychological assessments		33970
7	Psycholog$ test$		Psychology test, Psychological test, Psychology tests, Psychological tests		84204
8	5 or 6 or 7		To include any of the search terms relating to psychometrics		261901
9	Therap$		Therapy, Therapies, Therapeutic, Therapeutically		6660146
10	Psychotherap$		Psychotherapy, Psychotherapies, Psychotherapeutic, Psychotherapeutically		315450
11	Psycholog$ treatment$		Psychology treatment, psychology treatments, psychological treatment, psychological treatments		10495
12	9 or 10 or 11		To include any of the search terms relating to therapy		6791334
			To include:	To exclude:	
13	Psycholog$	Not drug$	Psychology, Psychological	Drug, Drugs	1751933
14		Not medication		Medication	1860361
15		Not surg$		Surgery, Surgical	1834010
16		Not child$		Child, Children, Childhood	1538817
17	13 and 14 and 15 and 16 and 17		To include only psychological studies meeting the eligibility criteria		1366307
18	4 and 8 and 12 and 17		To combine all of the key concepts in one search		1829
19	Limits applied		Adult, Human, English language, Peer-reviewed journal articles		886
20	Duplicates removed				756

Journal selection

A total of 756 journal articles were obtained using the above search strategy on 30th May 2014. An additional search strategy of offline sourcing (reference list trawling of relevant journal articles) was also employed, however no additional studies were obtained. Articles were screened using the method detailed in Figure 1.

Figure 1

Journal article selection process



In addition to the eligibility criteria produced for the database search, additional exclusion criteria were included when reviewing journal articles by title, abstract and full text, for example: exclusion of studies where ‘assessment’ referred to anything other than completing psychometric testing (e.g., initial interviews or file reviews).

Data abstraction

After extrapolating the main characteristics and key findings of all 16 studies, the methodological rigour of the studies was assessed. Given that it has been estimated that only 20% of literature is scientifically sound (Demaerschalk, 2004; Rychetnik & Wise, 2004) use of an effective method of evaluating studies and identifying those that have a rigourous methodology seemed particularly important. Therefore, the studies were reviewed for methodological quality using criteria developed by the first author (Table 2; Appendix A) on the basis of similar quality rating tools, such as the Critical Appraisal Skills Programme (CASP; Institute of Health Science, 1993) tool and the Newcastle-Ottawa Scale (NOS; Wells et al., 2010). We developed a new quality rating tool as we felt that currently available tools did not accurately analyse the quality of the papers for the question at hand. Ten criteria were developed, covering four broad areas of potential bias. Each criterion was rated 1-3 depending on its methodological quality. An overall score was achieved by summing these points together. Higher scores indicated a better quality paper. To allow comparison across quantitative, qualitative and mixed method papers, a percentage score was then calculated.

Results

Characteristics and assessed quality of retrieved studies

A summary of the characteristics of identified studies, and their key findings, is presented in Table 3 (Appendix B). Table 4 below provides an overview of the methodological quality of these studies, scored against the criteria outlined previously. The following sub-sections describe the results of the quality appraisal in greater detail, examining each criterion in turn.

Table 4

Quality appraisal

Study	Utility[1]	Quality criteria											Sum total	%
		1	2a	2b	3	4	5	6	7	8	9	10		
1	Y	**	**	**	**	**	**				*	*	14	58
2	P	**	**	**	**	**	**				**	*	15	63
3	P	**	***	**	***	***	***				***	**	21	88
4	Y	*	**	**	***	***	**				**	**	17	71
5	P	*	*	**	***	***	**				*	*	14	58
6	Y	** *	***	*	***	**	**	*	*	***	*	*	21	63
7	Y	**	***	*	**	**	**				**	*	15	63
8	Y	** *	***	*	***	***	***	***	***	***	***	***	25	93
9	P	*	***	*	***	***	**	*	*	**			17	63
10	Y	** *	***	***	***	**	**	***	***	***			25	93
11	P	**	***	***	*	***	***				***	**	20	83
12	Y	** *	***	**	***	***	**	**	***	***			24	89
13	Y	*	***	*	*	***	**				*	*	13	54
14	Y	*	***	**	***	***	**				*	**	17	71
15	P	*	***	**	***	***	**				*	**	17	71
16	Y	*	***	**	***	***	***	**	***	***			23	85

[1]Indicative of the utility of psychometrics; Y = Yes, only positive key findings extrapolated; P = Partially, key findings reported both positives and negatives of the utility of psychometrics for therapy; N = No, only negative key findings extrapolated. Percentage (%) calculated in order to compare quality criteria scores of all studies (as sum total possible for quantitative studies = 27, sum total possible for qualitative studies = 24, sum total possible for mixed method studies = 33).

Inclusion/ Exclusion criteria. A lack of defined inclusion and exclusion criteria poses limitations on the replicability of a study. It also makes it hard to generalise findings if no distinct sample has been identified, as different samples may react differently to the assessment process. Seven of the studies reviewed did not objectively define criteria for including or excluding participants in their study (Ackerman, Hilsenroth, Baity & Blagys, 2000; Ashworth et al., 2005; Costa & McCrae, 1992; Finn, 2011a; Finn, 2011b; Finn, 2012; Tiegreen, Braxton, Elbogen & Bradford, 2012) and a further five studies only partially specified their criteria, or criteria were indirectly reported (Cerney, 1983; Hare, Jones & Paine, 1999; Kelly et al., 2012; Mortimer & Smith, 1983; Ward, 2008).

Psychometric tests detailed. Different psychometric tests may have varying effects on the clients undertaking them, for example self-report measures could be deemed less stressful than those requiring an assessor, and this could then impact on the assessment process. However, if psychometric tests used in studies are not fully specified then it is difficult to have a comprehensive understanding of the subsequent observations of testing and to generalise and replicate findings. The psychometric tests were not specified in two out of the 16 studies reviewed (Tiegreen et al., 2012; Ward, 2008) and were only partially specified in three of them (Cerney, 1978; Hare et al., 1999; Mortimer & Smith, 1983).

From the 14 studies that provided partial/full information on the assessment measures used (Ackerman et al., 2000; Anker, Duncan & Sparks, 2009; Ashworth et al., 2005; Cerney, 1978; Costa & McCrae, 1992; Finn, 2011a; Finn 2011b; Finn 2012; Finn & Tonsager, 1992; Hare et al., 1999; Hilsenroth et al., 2004; Kelly et al., 2012; Mortimer & Smith, 1983; Schwartz, Merriman, Reed & Byock, 2005) 32 different psychometric tests were used. Twelve studies used more than one assessment with the remainder using only one, or not specifying the number of tests used.

Assessment process and feedback process. Examining only the assessment process or only the feedback process limits the ability of the research to holistically examine the impact of testing on the therapeutic process. Usefully, 11 studies examined this concept as a whole, or had findings extrapolated that

related to both areas (Ackerman et al., 2000; Ashworth et al., 2005; Costa & McCrae, 1992; Finn, 2011a; Finn, 2011b; Finn, 2012; Finn & Tonsager, 1992; Hilsenroth et al., 2004; Kelly et al., 2012; Tiegreen et al., 2012; Ward, 2008).

Impact on professional and client. The therapy process is often described as an interactive process between the professional and the client (e.g. see Stern et al., 1998) and it is therefore important to consider the impact of this process on both parties. However, only 4 of the 16 studies considered the impact of the use of tests on both the professional and the client (Ackerman et al., 2000; Finn & Tonsager, 1992; Kelly et al., 2012; Ward, 2008), with the remainder focusing on only one of these parties (mostly the clients rather than the impact on professionals).

Control group. Without the use of a comparison group it is difficult to establish whether findings are specific to a target population or not (Hermann & Whitman, 1984). Of the six quantitative studies in this review, only two had a comparison group that allowed reasonably specific conclusions to be drawn; Finn & Tonsager (1992), who compared those receiving test feedback (experimental group) to those who did not (control group) and Anker et al. (2009) who did similarly. Costa and McCrae (1992) and Schwartz et al. (2005) both did not have a comparison group in their research, although, Costa and McCrae's (1992) study was with a non-clinical sample. However, Schwartz et al.'s (2005) study examined a quality of life measure for palliative and end-of-life patients, where it may have been useful to see whether the impact varied with differing populations or with administration by different professionals.

The sample size also varied greatly between studies (sample size range: 1-205), with only five studies with more than 50 participants (Ackerman et al., 2000; Anker et al., 2009; Costa & McCrae, 1992; Finn & Tonsager, 1992; Schwartz et al., 2005). However, it should be noted that this is felt to be, in part, due to the inclusion of qualitative studies. Due to the number of qualitative studies (and in particular, case study designs) it is felt that this would be a plausible reason why there were a low number of studies that included a control group in their methodology.

Demographics and demographic matching. Demographic variables such as age, gender, number of years experiencing diagnosis/difficulty, social stability and socio-economic status are known to have an impact on how one experiences therapy and on therapy retention rates (e.g., Baekeland & Lundwall, 1975; Campbell, Darke & Popple, 2010; Fuller, 2009). Although none of the studies reviewed reported all of these demographics, 12 studies reported the age and gender of the clients (Ackerman et al., 2000; Anker et al., 2009; Costa & McCrae, 1992; Finn, 2011b; Finn, 2012; Finn & Tonsager, 1992; Hare et al., 1999; Hilsenroth et al., 2004; Kelly et al., 2012; Schwartz et al., 2005; Tiegreen et al., 2012; Ward, 2008) and two studies reported the same for the professionals involved (Anker et al., 2009; Ward, 2008). Descriptive data was only complete (i.e. for both the clients and the professionals) for two out of the 16 studies (Anker et al., 2009; Ward, 2008).

In view of the potential influence of demographic variables on the therapy process, it is encouraging to observe that, in all of the studies that used a comparison group, researchers attempted to match groups on demographic characteristics and test for any pre-existing between-group differences (i.e., potential confounds). Finn and Tonsager (1992) matched participants who completed the MMPI-2 and several outcome measures and received test feedback (experimental condition) with those in the “attention-only control group” who completed outcome measures, but did not receive any feedback. The control group did, however, receive “examiner attention” that the experimental group did not. It may have been useful in this particular study to have also had a group that completed the MMPI-2, but who did not receive feedback on it (to control for the fact that any differences may be down to the administration process of the MMPI-2, rather than the process of receiving feedback as the paper title suggests, and it would have been interesting to distinguish between the two).

The majority of studies focused on a clinical population sample, with only one study stating explicitly that they used a non-clinical sample (Costa and McCrae, 1992). The breadth of sampling across these studies was vast, ranging from a plethora of Axis I Disorders (e.g. Adjustment Disorder; Hilsenroth et al., 2004) through to Axis II Disorders (e.g. personality disorder not otherwise specified; Tiegreen et al., 2012) and neuropsychological deficits (not specified by the study; Ward, 2008). As a result of the breadth of disorders covered and the

lack of full demographics available, conclusions about this research area are hard to draw.

Standardised measures. The use of standardised measures is important as it means that firm conclusions can be drawn about the results with the assurance that scores were analysed through norm-referencing or criterion-referencing. The test scores can then be shown to have a degree of reliability and validity, as well as being generalisable and replicable (Kuncel & Hezlett, 2007). We would argue that the use of standardised measures means that the process of administration will also be somewhat standardised – such that conclusions about the *process* of administering a given test may transfer to other instances of administering that same test. In this review, four of the seven quantitative studies made use of standardised measures (Ackerman et al., 2000; Anker et al., 2009; Finn & Tonsager, 1992; Hilsenroth et al., 2004; Schwartz et al., 2005) and one used a standardised measure that was a novel measure (Costa & McCrae, 1992).

Rigour. The ability to understand the rationale for methods and forms of analysis undertaken is vital for the reader to be able to conceptualise the study, understand the research findings fully and trust them. Validity and reliability are easy ways of establishing this in quantitative research (Payton, 1979), but doing so for qualitative studies is harder. Sandelowski (1993) suggested that qualitative researchers generally unsuccessfully understand or achieve this. The rigour of the qualitative studies varied greatly; two were seen to provide a rich description and rationale of their work (Kelly et al., 2012; Ward, 2008), such as describing the decision-making process behind using a mixture of thematic analysis, an empirical phenomological method and a model for grounded research and summarising the steps involved during this process (Ward, 2008), three gave a limited description of this (Cerney, 1978, Finn, 2011a, Hare et al., 1999) and the remaining six gave little or no information at all.

Reflexivity. The degree to which researchers are reflexive in qualitative research greatly affects the degree to which a reader can make inferences about the data. If a researcher is not clear about, and does not reflect in-depth about, their own biases and assumptions then a reader cannot draw strong conclusions

about the findings. As Watt (2007) states, although there are guidelines to steer qualitative researchers, each piece of research is unique due to the researcher's individual stance and viewpoint, therefore reflexivity is essential in order to understand the research.

Unfortunately, a salient limitation of the qualitative studies in this review is that they lacked clear reflexivity. Five studies provided some (limited) evidence of taking a reflexive position (Finn, 2011a; Finn, 2011b; Finn, 2012; Kelly et al., 2012; Ward, 2008); for example, demonstrating an awareness of potential biases and broader influences on their approach (Finn, 2011a). However, six studies failed to address reflexivity at all (Ashworth et al., 2005; Cerney, 1978; Hare et al., 1999; Mortimer & Smith, 1983; Schwartz et al., 2005; Tiegreen et al., 2012). Furthermore, none of the papers stated the epistemological stance of the researcher(s).

Synthesis of findings

Perceived effects of the psychometric assessment process

Notwithstanding the methodological limitations outlined above, it was possible to extract findings to address our primary review question (Table 3; Appendix B): i.e., to determine the perceived effects of the psychometric assessment process (administration and feedback) on the therapist, the client, and the therapeutic process. The majority of findings were considered to be broadly positive – indicating beneficial contributions of the process – with the most common (consistently positive) finding being that psychometric assessment opened the way for dialogue and discussions (Ackerman et al., 2000; Anker et al., 2009; Ashworth et al., 2005; Cerney, 1978; Finn, 2011a; Finn, 2011b; Finn & Tonsager, 1992; Hilsenroth et al., 2004; Schwartz et al., 2005; Tiegreen et al., 2012). Some studies also identified negative findings – indicating a detrimental impact on therapy – with common issues including: (1) dependence on the capability of the professional (e.g., to interpret scores, or effectively communicate scores to the client; Costa & McCrae, 1992; Cerney, 1978; Ward, 2008); (2) concerns about the emotional response of the client (e.g., embarrassment, defensiveness, potential disappointment; Ashworth et al., 2005; Costa & McCrae, 1992; Ward, 2008) and (3) potential to adversely affect the therapeutic alliance/ cause ruptures (Ashworth et al., 2005; Kelly et al., 2012; Ward, 2008). With respect to the latter issue, it should be noted that most (ten) studies reported findings suggesting that the use of psychometric assessment helped to *strengthen* the therapeutic alliance (Ackerman et al., 2000; Anker et al., 2009; Costa & McCrae, 1992; Finn, 2011b; Finn & Tonsager, 1992; Hilsenroth et al., 2004; Tiegreen et al., 2012) – e.g., through fostering collaborative working and building rapport.

Analysing across retrieved studies, we synthesised individual findings into broader content-related categories: Discrete categories were formed for any semantically-linked findings that occurred across three or more papers (and which could not be adequately subsumed under other categories). These categories were grouped and organised in terms of implications for (1) the

therapy process, (2) the client, or (3) the therapist (whilst acknowledging that implications are somewhat interdependent). Table 5 summarises this synthesis.

Table 5

Main findings for perceived effects of psychometric assessment on therapy process, client, and therapist

	Article															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Therapy Process																
Structure/focus	+		±		-	+		+		+	+				+	
Alliance			±		-	+		+	+	+	±	+	+		+	+
Communication				+	+	+		+		+		+	+	+	+	+
Client																
Validation			+	+			+	+			+	+				+
Empowerment			±		+						+		+			
Insight					+			+	+		+		+	+	+	+
Hope/expectancy								+		+			+			
Emotion					-				-		-	+				
Engagement and retention					+			+		+	+		+			+
Therapist																
Understanding		+											+	+	+	
Empathy				+					+			+	+			
Care planning						+									+	+
Working within capabilities		-							-		-					

+ denotes a perceived positive effect; - denotes a perceived negative effect; ± denotes both positive and negative (i.e., mixed) effects.

As can be seen from Table 5, synthesised findings suggested that the greatest impact of the testing process on clients was in helping them to feel understood (*validation*) and to conceptualise their difficulties (gain *insight*). For therapists/professionals, psychometric test use was found to have beneficial impact through facilitating *empathy* and *understanding.* As discussed above, synthesised findings supported beneficial impact on therapy process in terms of strengthening *communication* and *alliance* (although findings also identified potential for negative impact on the latter aspect of therapy process). Another notable perceived benefit of psychometric assessment/feedback was in bringing *structure/ focus* to therapy, although two studies identified that assessment administration could be seen as 'intrusive' and incongruent with the therapy

process (i.e., potentially detrimental to *structure/ focus*, depending on how and when assessment is introduced).

Conclusion

This review examined whether the process of psychometric assessment might have some utility for therapy beyond the functional production of information (e.g., scores to evaluate treatment outcomes). The aim was to understand the impact of the act of the assessment process (completing psychometric assessment(s) and assessment feedback, if given) on the therapeutic process, the client and the professional. A mixed pattern of results emerged, with just under two-thirds of the studies providing evidence to suggest the positive effects of the assessment process and just over one-third of the studies providing evidence depicting both benefits and drawbacks. As much of the evidence had to be extrapolated, due to only a minority of studies actually focusing on the review question at hand, it seems that this area of research is still in its infancy. Moreover, these studies evinced multiple methodological flaws that limit our confidence in reported results. These ranged from issues with clarity of sample characteristics to a lack of explicit researcher reflexivity. Qualitative research generally failed to manage the threats of trustworthiness discussed by Padgett (1998), such as reflexivity and rigour.

Although this review adopted a rigourous methodology itself in an attempt to reduce risk of bias, this review also has limitations; for example, search terms may have been too narrow and exclusion criteria too restrictive. There are also the wider concerns of the validity and reliability of the test measures that we as clinicians use, problems with malingering that might be experienced, location of test administration, the qualities of the test administrator, the client's past experience or cultural expectation of psychometrics assessments, the impact and interaction of different disorders and the way feedback is communicated (if given at all), none of which were addressed by this review.

However, it is important to bear in mind that despite the mixed conclusion as to the effects of the psychometric assessment process given the methodological weaknesses of the studies involved in the review, an absence of clear results does not mean the absence of an 'effect' (Altman & Bland, 1995). Emergent themes of opening the way for dialogue, establishing a focus and a framework for therapy and encouraging/ motivating the client are positive indications from the research. In terms of future research, studies would do well

to look into these emergent themes to provide them with more substantial backing. Potential training (or re-training) of professionals may also be an important future implication with regards to the concerns raised in some of the reviewed papers over the ability of the professional during the assessment process.

Reflexivity on the process

We felt that it was important to employ our own critical appraisal criteria to this systematic literature review and have therefore provided a brief reflection on our biases/ assumptions in undertaking this piece of research. We acknowledge that we have a personal interest in neuropsychology and the use of psychometric tests, with two of the authors working within this sub-speciality of clinical psychology. We were therefore potentially biased in that we hoped that the effects of the psychometric assessment process would be beneficial. However, we were open about our biases, meaning we could be aware of their impact at all stages of the research. Extracting data from other studies allowed us to retain a more neutral stance throughout this process and synthesising individual findings into broader content-related categories of positive and negative perceived effects, and cross-referencing these as authors, also helped us to check reliability and to not make assumptions about the data. As a result, we were able to report what we feel is a balanced discussion of both positives *and* identified negatives from the literature.

References

** by a reference denotes that it is one of the 16 reviewed journal articles.*

* Ackerman, S. J., Hilsenroth, M. J., Baity, M. R., & Blagys, M. D. (2000). Interaction of therapeutic process and alliance during psychological assessment. *Journal of Personality Assessment, 75(1),* 82-109.

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *BMJ*, *311*, 485.

* Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trail in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77(4),* 693-704.

* Ashworth, M., Robinson, S. I., Godfrey, E., Parmentier, H., Sheprherd, M., Christery, J., Wright, K., & Matthews, V. (2005). *Counselling and Psychotherapy Research, 5(1),* 37-42.

Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., et al. (2004). A client-centred psychometric instrument: The development of 'PSYCHLOPS' ('Psychological Outcomes Profiles'). *Counselling and Psychotherapy Research, 4,* 27-33.

Baekeland, F. & Lundwall, L. (1975). Dropping out of treatment. A critical review. *Psychological Bulletin, 82(5),* 738-783.

Barber, J. P., & Crits-Christoph, P. (1996). Development of a therapist adherence/competence rating scale for Supportive/Expressive Dynamic Psychotherapy: A preliminary report. *Psychotherapy Research, 6,* 79–90.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Campbell, G., Darke, S., & Popple, G. (2010). *Effects of client characteristics and mental health on treatment completion and retention in a therapeutic community.* Sydney: National Drug and Alcohol Research Centre.

Cella, D. F. S., Jacobsen, P. B., Holland, J. C., Orav, J., Silberfarb, P. M., & Rafla, S. (1987). A brief POMS measure of distress for cancer patients. *Journal of Chronic Diseases, 40,* 939-942.

* Cerney, M. S. (1978). Use of the psychological test report in the course of psychotherapy. *Journal of Personality Assessment, 42(5),* 457-463.

Cheek, J. M., & Buss, A. H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology, 41,* 330-339.

Clemence, A., Hilsenroth, M., Ackerman, S., Strassle, C., & Handler, L. (2005). Facets of the therapeutic alliance and perceived progress in psychotherapy: relationship between patient and therapist perspectives. *Clinical Psychology & Psychotherapy, 12(6),* 443-454.

Costa, P. T., Jr., & McCrae, R. R. (1989). *The NEO-PI/NEO-FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practise: The NEO Personality Inventory. *Psychological Assessment, 4(1),* 5-13.

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan Executive Function System*. San Antonio, TX: Psychological Corporation.

Diamond, G. S., Liddle, H. A., Wintersteen, M. B., Dennis, M. L., Godley, S. H., & Tims, F. (2006). Early therapeutic alliance as a predictor of treatment outcome for adolescent cannabis users in outpatient treatment. *American Journal on Addictions, 15(1),* 26-33.

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., et al. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health, 9,* 247-255.

Exner, J. (2003). *The Rorschach: A comprehensive system* (4th ed.). New York, NY: Wiley.

Fechner, G.T. (1860). *Elemente der Psychophysik.* Leipzig: Breitkopf und Härtel. In Fechner, G.T. (1966). Elements of Psychophysics. (Vol. 1). (H. E. Adler, Trans.). New York: Holt, Rinehart and Winston.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43,* 522-527.

* Finn, S. E. (2011a). Use of the Adult Attachment Projective Picture System (AAP) in the middle of a long-term psychotherapy. *Journal of Personality Assessment, 95(5),* 427-433.

* Finn, S. E. (2011b). Journeys through the valley of death: Multimethod psychological assessment and personality transformation in long-term psychotherapy. *Journal of Personality Assessment, 93(2),* 123-141.

* Finn, S. E. (2012). Implications of recent research in neurobiology for psychological assessment. *Journal of Personality Assessment, 94(5),* 440-449.

* Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4(3),* 278-287.

Finn, S. E. & Tonsager, M. E. (1997). Information-gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment, 9,* 374-385.

Fischer, C. T. (1994). *Individualised psychological assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Folstein, M. F. (1983). The min-mental status examination. In Crook, T., Ferris, S., Bartus, R. (eds.). *Assessment in geriatric psychopharmacology* (pp50-51). New Canaan, CT: Mark Powley.

Fuller, S. E. (2009). Pretreatment client characteristics and treatment retention in an intensive outpatient substance abuse treatment program. *Dissertations, Paper 1.* Retrieved August 9, 2013, from http://epublications.marquette.edu/dissertations_mu/1

Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales-Patient Version. *Psychological Assessment, 3,* 68-74.

George, C., & West, M. (2001). The development and preliminary validation of a new measure of adult attachment: The Adult Attachment Projective. *Attachment and Human Development, 3,* 30-61.

Hare, D. J. (1997). The use of repertory grid techniques in the assessment of people with learning disabilities: A case study. *Journal of Learning Disabilities for Nursing, Health and Social Care, 1(3),* 115-120.

* Hare, D. J., Jones, J. P. R., & Paine, C. (1999). Approaching reality: The use of Personal Construct Assessment in working with people with Asperger Syndrome. *The National Autistic Society, 3(2),* 165-176.

Hatcher, R. L. (1999). Therapists' views of treatment alliance and collaboration in therapy. *Psychotherapy Research, 9,* 405-423.

Hatcher, R. L., & Barends, A. W. (1996). Patient's view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measure. *Journal of Consulting and clinical Psychology, 64,* 1326-1336.

Hermann, B. P., & Whitman, S. (1984). Behavioral and personality correlates of epilepsy: A review, methodological critique, and conceptual model. *Psychological Bulletin*, *95*(3), 451-497.

* Hilsenroth, M. J., Peters, E. J., & Ackerman, S. J. (2004). The development of therapeutic alliance during psychological assessment: Patient and therapist perspectives across treatment. *Journal of Personality Assessment, 83(3),* 332-344.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56,* 885–892.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology, 36,* 223-233.

Horvath, A. O. & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38(2),* 139-149.

Jackson, D. N. (1989). *Basic Personality Inventory manual.* Port Huron, MI: Sigma Assessment Systems.

* Kelly, V., Holttum, S., Evans, C., & Shepherd, M. (2012). A discourse analysis of power in relation to PSYCHLOPS (Psychological outcome profiles) in the context of CT for psychosis. *Counselling and Psychotherapy Research: Linking research with practise, 12(4),* 247-256.

Kuncel, N. R., & Hezlett, S. A. (2007). *Science, 315*, 1080-81.

Kuyken, W. & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*, 42–48.

Miller, S. D., & Duncan, B. L. (2004). *The Outcome and Session Rating Scales: Administration and scoring manual.* Fort Lauderdale, FL: Author.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet, 354(9193*), 1896-1900.

Morey, L. (1991). *Personality Assessment Inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.

* Mortimer, R. L. & Smith, W. H. (1983). The use of the psychological test report in setting the focus of psychotherapy. *Journal of Personality Assessment, 47(2),* 134-138.

Murray, H. A. (1943). *Thematic Apperception Test manual.* Cambridge, MA: Harvard University Press.

NICE. (2009). *Depression: the treatment and management of depression in adults (update).* Retrieved August 7, 2013, from http://www.nice.org.uk/CG90

OvidSP. (2013). Retrieved August 3, 2013, from http://ovidsp.tx.ovid.com.ezproxy.nottingham.ac.uk/sp-3.4.1b/ovidweb.cgi .

Oxford Dictionary. (2013). Retrieved August 6, 2013, from http://oxforddictionaries.com/definition/english/psychometrics .

Padgett, D. K. (1998). *Qualitative Methods in Social Work Research: Challenges and Rewards.* Thousand Oaks, CA: Sage.

Portenoy, R. K., Thaler, H. T., Kornblith, A. B., Lepore, J. M., Friedlander-Klar, H., Kiyasu, E., Sobel, K., Coyle, N., Kemeny, N., Norton, L. et al. (1994). The Memorial Symptom Assessment Scale: An instrument for the evaluation of symptom prevalence, characteristics, and distress. *European Journal of Cancer, 30A,* 1326-1336.

Rapaport, D., Gill, M., & Schafer, R. (1968). Diagnostic psychological testing. In R. R. Holt (Ed.), New York: International Universities Press.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069-1081.

Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advance in Nursing Science, 16(2),* 1-8.

* Schwartz, C. E., Merriman, M. P., Reed, G., & Byock, I. (2005). Evaluation of the Missoula-VITAS Quality of Life Index-Revised: Research tool or clinical tool? *Journal of Palliative Medicine, 8(1),* 121-135.

Sexton, H., Littauer, H., Sexton, A., & Tommeras, E. (2005). Building an alliance: Early therapy process and the client-therapist connection. *Psychotherapy Research, 15(1-2),* 103-116.

Stern, D. N., Sander, L. W., Nahum, J. P., Harrison, A. M., Lyons-Ruth, K., Morgan, A. C., Bruschweilerstern, N., & Tronick, E. Z. (1998). Non-interpretative mechanisms in psychoanalytic therapy: The 'something more' then interpretation. *International Journal of Psycho-Analysis, 79,* 903-921.

Stiles, W. B.,&Snow, J. S. (1984). Counseling session impact as viewed by novice counselors and their clients. *Journal of Counseling Psychology, 31,* 3–12.

Tharinger, D. J., Finn, S. E., Hersh, B., Wilkinson, A., Christopher, G. B., & Tran, A. (2008). Assessment feedback with parents and preadolescent children: A collaborative approach. *Professional Psychology: Research and Practise, 39(6),* 600-609.

* Tiegreen, J. A., Braxton, L. E., Elbogen, E. B., & Bradford, D. (2012). Building a bridge of trust: Collaborative assessment with a person with serious mental illness. *Journal of Personality Assessment, 94(5),* 513-521.

* Ward, R. M. (2008). Assessee and assessor experiences of significant events in psychological assessment feedback. *Journal of Personality Assessment, 90(4),* 307-322.

Watt, D. (2007). On becoming a qualitative researcher: The value of reflexivity. *The Qualitative Report, 12(1),* 82-101.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV).* San Antonio, TX: NCS Pearson.

Weiss, L. (2004). *Therapist's Guide to Self-Care.* Hove, East Sussex: Brunner-Routledge.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry, 33,* 1111–1115.

Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2010). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Retrieved August 3, 2013, from http://www.ohri.ca/programs/clinical _epidemiology /oxford.asp,

Appendices

Appendix A. Quality appraisal tool

Table 2

Quality criteria

		Quality control	Rating
Participants	1	Inclusion and exclusion criteria specified	*** Both inclusion and exclusion criteria are clearly reported; ** Inclusion and exclusion criteria are partially or indirectly reported; * No inclusion or exclusion criteria are reported.
	2a	Participant demographics of client(s) included	*** Demographics of the client(s) are clearly reported, including age and gender; ** Demographics of the client(s) are partially reported; * Demographics of the client(s) are not adequately reported or are not present.
	2b	Participant demographics of professional(s) included	*** Demographics of the professional(s) conducting the assessment(s) are clearly reported, including age, gender and profession; ** Demographics of the professional(s) are partially reported; * Demographics of the professional(s) are not adequately reported or are not present.
Other	3	Psychometric tests used were detailed for the reader	*** All psychometric tests used in the assessment conducted for the research are clearly reported; ** Psychometric tests used in the assessment are partially reported; * Psychometric tests used in the assessment are not adequately reported (e.g.; used an assessment battery) or are not stated at all.
	4	Examined process of assessment and assessment feedback	*** The research examined both the process of assessment and the impact of assessment feedback; ** The research examined only the process of assessment or only the process of assessment feedback; * The research was not adequately clear about what process was being examined.
	5	Examined process from the viewpoint of the client(s) and the professional(s)	*** The viewpoint of both the client(s) and the professional(s) was clearly reported; ** The viewpoint of both the client(s) and the professional(s) was moderately or partially reported; * The viewpoint of only the client(s) or the professional(s) was reported.
Quantitative issues	6	Use of a control/comparison group	*** There was a comparison group allowing reasonably specific conclusions to be drawn; ** There was a control group but it only allowed general conclusions to be drawn; * There was no control group or their data was not analysed.
	7	Groups demographically matched (* awarded in the absence of a control/comparison group)	*** Reported demographic variables were matched; ** Reported demographic variables were un-matched or data was reported without statistical comparison; * The groups differed in several ways that were not statistically corrected or there was no data.
	8	Standardised measures used	*** Appropriate and standardised measures were used; ** Appropriate but adapted or modified measures are used or measures are novel; * No standardised measures are used.
Qualitative issues	9	Rich rigour	*** The study provides a rich description and rationale for the methods and forms of analysis undertaken; ** The study provides a less detailed or limited description and rationale for these criteria; * Little or no information is given to be able to adequately assess these criteria.
	10	Reflexivity of the researcher(s)	*** The researcher(s) are transparent about their subjective biases, provide sufficient detail on them and reflect upon the impact of these on the research; ** The researcher(s) provide less detailed description of these criteria but do attempt to give some insight into researcher assumptions; * Little or no information is apparent to adequately assess these criteria.

Appendix B. Characteristics and key findings of retrieved studies

Table 3

General characteristics of retrieved studies

First author (date)	Methodology / Perspective	Sample	Focal assessment measure(s)	Key finding(s) relating to review[1]
1. Mortimer (1983)	Qualitative (Case study design) / Professional	**Professional(s)** Psychotherapist n=2 Gender and age unknown **Client(s)** Male n=1 Female n=2 **Clinical population** Not specified	**Main focus** Assessment modelled after Rapaport's test battery, inc. TAT	Use of a test report helps to: 1. establish a central focus at the initial stages of therapy and 2. regain focus when reviewed. By establishing a clear focus/direction of therapy changes occur.
2. Cerney (1978)	Qualitative (Case study design) / Professional	**Professional(s)** Psychotherapist n=1 Gender and age unknown **Client(s)** Female n=3 Age unknown n=1 Age range=20-30 n=2 **Clinical population** Not specified	**Main focus** Not specified **Other** TAT	By re-reading test reports at varying stages of treatment one can remain objective about the client Acts as a catalyst to exploring underlying feelings of client *Concerns about the capability of the psychologist to: 1. make inferences about the data and 2. communicate the results effectively to the client.*
3. Kelly (2012)	Qualitative (Foucauldian discourse analysis) / Professional & Client	**Professional(s)** Clinical psychologist n=3 CBT therapist n=1 Age range=26-45 **Client(s)** Female n=1 Male n=3 Age range=26-45 **Clinical population** Psychosis (schizophrenia, schizoaffective disorder, psychotic depression)	**Main focus** CORE-OM PSYCHLOPS	Supports collaborative working Focuses discussion, helps prioritise client's difficulties Legitimises client's difficulties Emphasised power of the client Client felt contained by assessment process *Oppressive for client – just ticking boxes* *Feels invasive to therapy* *If assessment done too early can adversely affect: 1. engagement and 2. the therapeutic alliance.* *Free text options not containing*

4. Finn (2011a)	Qualitative (Single case design) / Professional	**Professional(s)** Clinical psychologist n=1 Male n=1 Age unknown **Client(s)** Male n=1 Age "late 40s" **Clinical population** Not specified. Client referred initially for Couples Therapy.	**Main focus** AAP **Others** MMPI-2 Rorschach	Keeping test results in mind increases therapist patience AAP training increases therapist empathy Catalyst to accessing difficult emotions and stimulating discussion, particularly when test report shared with client Client feels: 1. understood and 2. reassured.
5. Ashworth (2005)	Qualitative / Professional	**Professional(s)** Clinical psychologist n=1 Counsellor n=2 Psychotherapist n=1 Males n=2 Females n=2 Age unknown **Clinical population** Not specified	**Main focus** CORE-OM PSYCHLOPS	Helps clients conceptualise their problems Allows differences in therapist's and client's perceptions to be explored Empowers and encourages client (to see pre- post-changes) Helps therapy retention rates (as therapist understands client before therapy even embarks) *Intrusion into therapy time impacting negatively on:* *1. therapy outcome and* *2. therapeutic alliance.* *Client embarrassment (over expression, handwriting, literacy)*
6. Schwartz (2005)	Mixed methods / Professional	**Professional(s)** Staff working in hospice, home health and palliative care Gender and age unknown **Client(s)** Male n=98 Female n=67 Mean age=66.3 **Clinical population** Palliative and end-of-life (end-care renal patients on dialysis, hospice	**Main focus** MVQOLI-R **Others** Brief POMS Folstein Mini-mental Status Exam MSAS MVQOLI Ryff Psychological Well-Being measure – short form	Opened the door for discussion, provided framework and language for (awkward) issues Facilitated holistic, collaborative care Useful for care planning

		or long-term care patients)		
7. Hare (1999)	Qualitative / Professional	**Professional(s)** "Therapist" Gender and age unknown **Client(s)** Male n=4 Age range=21-35 **Clinical population** Asperger syndrome	**Main focus** Personal Construct Assessment	Clients felt taken seriously (due to the length and apparent complexity of the assessment)
8. Finn (1992)	Quantitative (2x3 repeated-measures design) / Professional & Client	**Professional(s)** Not specified **Client(s)** Female n=42 Male n=18 Mean age=23.3 **Clinical population** Not specified. Clients were known not to be suicidal, psychotic or in danger of causing harm to self or others.	**Main focus** MMPI-2 **Others** SEQ(a) SCL-90-R SCI AQ	Process of sharing test results: 1. Builds rapport 2. Increases client cooperation 3. Leaves clients feeling positive about psychological testing and mental health professionals 4. Is therapeutic (if therapist actively encourages participation in the feedback session) Client feels: 1. Sense of relief 2. Understood, legitimate 3. Increased self-esteem 4. Increased hope 5. Increased self-awareness and understanding 6. Increased motivation (to seek mental health services or actively participate in on-going therapy) 7. Reduced isolation 8. Reduced symptomatic distress Written feedback report opens way for discussion and opportunities for re-framing
9. Costa (1992)	Quantitative / Professional	**Professional(s)** Not specified **Client(s)** Male n=60 Female n=57 Age range=21-94 Mean age= 67.5(m) 64.5(f) **Clinical population** Non-clinical	**Main focus** NEO-PI **Others** BPI PAI	More rapid development of rapport Therapists appear more knowledgeable and empathic Routine sharing of scores and referring back to them throughout therapy helps to achieve insight *Could increase client defensiveness* *Reaction of client dependent on therapist's ability to elicit the*

				client's trust, interest and cooperation *Confusion on the part of the therapist, e.g. if therapist unsure how to use the test information, if test seen to give 'false' information about clients potentially due to issues with social desirability, malingering or effects of psychopathology, e.g. narcissism*
10. Anker (2009)	Quantitative (2x3 repeated-measures design) / Professional	**Professional(s)** Licensed psychologist n=4 Licensed social worker n=5 Licensed psychiatric nurse n=1 Male n=3 Females n=7 Mean age=42 **Client(s)** 205 heterosexual couples Age range=20-71 Mean age=37.83 **Clinical population** Not specified	**Main focus** ORS LW Marital Adjustment Test SRS	Allows session focus Allows therapist to openly discuss any concerns Increases: 1. Retention (doubled) 2. treatment effect size 3. Engagement Monitoring throughout the therapeutic process can: 1. enhance client expectancy 2. amplify participation/efforts (due to sensitisation to the experience of change) 3. secure a strong alliance Individualises therapy
11. Ward (2008)	Qualitative (mixed methods: thematic analysis with features of an empirical phenomenolog-ical method and grounded theory) / Professional & Client	**Professional(s)** Clinical psychology final year intern n=3 Unlicensed doctoral fellow n=1 Licensed psychologist n=2 Male n=2 Female n=4 Age range=26-39 **Client(s)** Male n=3 Female n=3 Age range=19-26 **Clinical population** Not specified. Clients were "seeking psychological assessment for different reasons", either	Not specified	Increases collaboration Aids clients: 1. self-verification 2. self-enhancement 3. self-efficacy 4. self-discovery Interactive feedback process makes the process more individualised Client feels: 1. uniquely understood and more motivated (due to the individualised process) 2. more objective (due to comparisons with others helping one to engage in a less harsh self-evaluation) 3. Able to make sense of their difficulties prior to commencement of therapy (thus increasing therapeutic engagement)

		neuropsychological or academic concerns		Is transformative. Starts the process of change (from a view of global self-blame to one of informed action and autonomy) Provides: 1. an action plan and an agenda for therapy 2. a framework to develop meaningful and explanatory connections to the clients life *Client experiences:* *1. a degree of discomfort with the accuracy of results* *2. betrayal (that the assessor used information that they thought was 'off the record')* *3. emotional difficulty when told of findings they consider negative* *Creates ruptures and decreases engagement* *Therapist worries about their own ability as a therapist, resulting from concerns about:* *1. the effect of feedback on the clients* *2. the difficulty providing feedback on emotional functioning* *3. the challenge of providing an alternative explanation of the clients difficulties to the one they already have* *4. interpretative error (due to complexity of the findings)*
12. Hilsenroth (2010)	Quantitative (repeated-measures design) / Professional	**Professional(s)** Advanced doctoral students n=18 PhD licensed psychologist n=1 Female n=11 Male n=8 Age unknown **Client(s)** Female n=28 Male n=14 Mean age=30.6 **Clinical population** All had a primary DSM-IV Axis I diagnosis (Adjustment	**Main focus** CASF-P CASF-T **Others** WAI CALPAS TCC	Facilitates an empathic connection between therapist and client Increases collaboration Test feedback increases therapeutic alliance Opens opportunity for discussion Client feels: 1. Relief 2. Accepted/understood TMA approach enhances alliance over use of a traditional IG model of assessment

		disorder, Anxiety disorder, Eating disorder, Mood disorder, Substance-related disorder, V Code relational problem)		
13. Tiegreen (2012)	Qualitative (Single case design) / Professional	**Professional(s)** Not specified **Client(s)** Male n=1 Age=29 **Clinical population** Diagnoses of schizoaffective disorder, depressed type, alcohol abuse and personality disorder not otherwise specified (with paranoid and borderline features)	Not specified	Opens dialogue for discussions Provides opportunities to start the therapeutic process (psycho-education, normalisation) Builds rapport as results seen as more personally relevant Helps clients discover meaning and fresh perspectives/insight and empowers them to change Increases therapeutic alliance Increases collaboration Increases therapist: 1. Empathy towards the client 2. Opportunity to understand from client's perspective Increases client: 1. Motivation 2. Self-verification 3. Insight 4. Self-awareness 5. Self-discovery 6. Self-enhancement 7. Hope for ability to cope Referring back to results throughout process validates client and reassures them that they have been heard
14. Finn (2012)	Qualitative (Single case design) / Professional	**Professional(s)** Clinical psychologist n=1 Male n=1 Age unknown **Client(s)** Male n=1 Age=27 **Clinical population** Sexual addiction	**Main focus** MMPI-2 Rorschach **Other** TAT	Increases insight of therapist and client. This then helps to change the clients' view of themselves. Active client participation in assessments and assessment feedback sessions helps increase: 1. emotional impact 2. in-depth discussions
15. Finn (2011b)	Qualitative (single case design) /	**Professional(s)** Clinical psychologist n=1 Male n=1	**Main focus** MMPI-2 Rorschach	Mobilises powerful transference and counter-transference feelings useful for therapy

	Professional	Age unknown **Client(s)** Male n=1 Age=23 **Clinical population** Not fully specified. Problems with intimate relationships, depression.		Deepens rapport Guides and reassures therapists and clients Increases therapist's: 1. empathy 2. ability to make more realistic treatment recommendations 3. ability to attenuate to psychological processes during therapy 4. ability to make more fine-tuned interpretations Reduces therapist confusion over client presentation Use of the assessment results throughout the therapeutic process: 1. opens way for discussions 2. increases client insight 3. helps the therapist to maintain the correct pace *Use of assessment with the clients own therapist (as opposed to a separate assessor) can 'cloud' results*
16. Ackerman (2000)	Quantitative / Professional & Client	**Professional(s)** Advanced doctoral students n=18 Male n=9 Female n=19 Age unknown **Client(s)** Male n=54 Female n=74 Mean age=27.9 **Clinical population** Majority DSM-IV Axis I diagnoses (Academic disorder, adjustment disorder, anxiety disorder, ADHD, conduct disorder, eating disorder, identity problem, mood disorder, ODD, PTSD, substance-related disorder, V Code problem)	**Main focus** SEQ(b) CASF HAq-R Rorschach Social Adjustment Scale SCL-90-R PAI Inventory of Interpersonal Problems **Other** See paper	Opens dialogue Feels collaborative Helpful to develop a treatment frame Therapeutic assessment (in-depth assessment process) increases retention Facilitates client change Client feels: 1. understood 2. validated in their reality 3. increase in self-concept 4. more objective when examining their difficulties Good test feedback led to a good therapeutic alliance Length and 'smoothness' of feedback process found not to be as important, i.e. if level of depth reached in assessment then some discomfort over non-pleasurable results reported to be acceptable

	TMA reduces termination of therapy over a traditional IG approach

[1]Key findings are ideas extrapolated from the papers that relate to some aspect of the review question; Positive findings = normal text, Negative findings = italicized; Adult Attachment Projective Picture System (AAP; George & West, 2001); Assessment Questionnaire (AQ; Finn & Tonsager, 1992); Attention Deficit Hyperactivity Disorder (ADHD); Basic Personality Inventory (BPI; Jackson, 1989); Brief Profile of Mood States (Brief POMS; Cella, Jacobsen, Holland, Orav, Silberfarb & Rafla, 1987); California Psychotherapy Alliance Scale (CALPAS; Gaston, 1991); Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2000); Combined Alliance Short Form (CASF; Hatcher & Barends, 1996); Combined Alliance Short Form-Patient Version (CASF-P; Hatcher & Barends, 1996); Combined Alliance Short Form-Therapist Version (CASF-T; Hatcher, 1999); Diagnostic and Statistical Manual of Mental Disorders-fourth edition (DSM-IV; American Psychiatric Association, 2000); Folstein Mini-mental Status Exam (Folstein, 1983); Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureno & Villasenor, 1998); Locke Wallace Marital Adjustment Scale (LW Marital Adjustment Scale; Locke &Wallace, 1959); Memorial Symptom Assessment Scale (MSAS; Portenoy et al., 1994); Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen & Kaemmer, 1989); Missoula-VITAS Quality of Life Index (MVQOLI; Byock & Merriman, 1998); Missoula-VITAS Quality of Life Index-Revised (MVQOLI-R; Schwartz, Merriman, Reed & Byock, 2005); NEO Personality Inventory (NEO-PI; Costa & McCrae, 1989); Oppositional Defiant Disorder (ODD); Outcome Rating Scale (ORS; Miller & Duncan, 2004); Penn Helping Alliance Questionnaire-Revised (HAq-R; Barber & Crits-Cristoph, 1996); Personal Construct Assessment (Hare, 1997); Post Traumatic Stress Disorder (PTSD); Psychological Outcome Profiles (PSYCHLOPS; Ashworth et al., 2004); Rapaport's test battery (Rapoport, Gill & Schafer, 1968); Rorschach (Exner, 2003); Ryff Psychological Well-Being measure – short form (Ryff, 1989); Self-Consciousness Inventory (SCI; Fenigstein, Scheier & Buss, 1975); Self Esteem Questionnaire (SEQ(a); Cheek & Buss, 1981); Self Evaluation Questionnaire (SEQ(b); Stiles & Snow, 1984); Session Rating Scale (SRS; Miller & Duncan, 2004); Social Adjustment Scale (Weissman & Bothwell, 1976); Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1983); Thematic Apperception Test (TAT; Murray, 1943); Therapist Confident Collaboration Scale (TCC; Clemence, Hilsenroth, Ackerman, Strassle & Handler, 2004); Therapeutic Model of Assessment (TMA; Finn & Tonsager, 1992, 1997; Fischer, 1994); Traditional Information Gathering model of assessment (IG), Working Alliance Inventory (WAI; Horvath & Greenberg, 1989).

Journal Paper

A study of the normative psychometric properties of the ALFIE, a novel measure of accelerated long-term forgetting in Temporal Lobe Epilepsy.

Emma Cameron[a]

Nima Moghaddam[ab]

Roshan das Nair[cd]

[a]Trent Doctorate in Clinical Psychology, School of Life Sciences, Bridge House, Lincoln University, Brayford Campus, Lincoln, LN6 7DL, United Kingdom

[b]Corresponding author: Email: NMoghaddam@post01.lincoln.ac.uk; Telephone: +44 (0) 1522 886 029

[c]Trent Doctorate in Clinical Psychology, Division of Psychiatry and Applied Psychology, School of Medicine, YANG Fujia Building, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, United Kingdom

[d]Dept of Clinical Psychology and Neuropsychology, Nottingham University Hospitals NHS Trust, Nottingham NG7 2UH, United Kingdom

Abstract

Background: Patients with Temporal Lobe Epilepsy (TLE) typically complain of memory difficulties, but these are not always evident on objective memory assessments. This discrepancy may be caused by a lack of specificity in current memory measures that utilise delay periods of only 30-40 minutes. Research suggests that extending published memory tests to assess recall over a longer period (e.g., a two-week delay) allows these memory deficits to be seen – a phenomenon known as Accelerated Long-term Forgetting (ALF). However, the literature has yet to address the difficulties with test veridicality and clinical practicability, something this study aimed to examine through further development of a novel measure of ALF.

Method: 50 healthy participants' objective memory performance was assessed by asking them to recall and recognise information at three time points: immediately after presentation of a stimulus (T1), after 40 minutes (T2), and after two-weeks (T3). Alongside a published story and word list tasks, a novel measure was used – the Accelerated Long-term Forgetting In Epilepsy (ALFIE) test. This test uses multi-modal stimuli drawn from real-life televised news broadcasts, with telephone follow-up phone-calls to assess two-week recall and recognition. We believe this to be more ecologically valid and clinically practical then other memory measures. Subjective memory performance was also assessed via use of a self-report questionnaire.

Results: Convergence of results on the ALFIE and published memory measures was found. Although extension of the published measure to a two-week delay might then seem justifiable, the ALFIE test showed greater correlations with subjective memory scores than the published story and word lists tasks. This suggests greater ecological validity of the ALFIE measure than the published memory test. Reliability was assessed through inter-rater reliability and analysis of parallel forms. The ALFIE showed high inter-rater reliability and although parallel forms reliability was poor, through standardisation versions may be used as alternate forms. Low attrition rates suggest that use of a two-week delayed assessment via telephone might be a clinically viable solution for specialist

services assessing ALF, often over large geographical regions. Significant differences in performance between genders needs further examination, but may partially be explained through emotional salience of materials.

Conclusion: The ALFIE test appears to be a viable test measure for assessing memory that is more ecologically-valid and clinically practicable than current memory measures.

Keywords: Temporal Lobe Epilepsy, Accelerated Long-term Forgetting, Neuropsychological Assessment, Ecological Validity.

1. Introduction

1.1 Epilepsy and forgetting

Epilepsy is defined by the International League Against Epilepsy as a disease typified by recurrent seizures due to abnormal and excessive electrical activity in the brain (Fisher et al., 2014). Neuropsychological deficits are common in this population, with memory difficulties being the most common complaint (see e.g., Hendriks, Aldenkamp, Van der Vlugt, Alpherts & Vermeulen, 2002; Ponds & Hendriks, 2006; Thompson, & Corcoran, 1992) (see Extended paper 1.1). Studies have long suggested that the focus of seizures to the temporal lobe is a key contributing factor to memory impairment (Delaney, Rosen, Mattson, & Novelly, 1980; Giovagnoli & Avanzini, 1999; Viskontas, McAndrews, & Moscovitch, 2000). Indeed, those with Temporal Lobe Epilepsy (TLE), where seizures are localised to the temporal lobes of the brain, present with a specific memory impairment, whereby information is acquired and retained normally over delays of minutes/hours, but is characterised by an abnormally fast rate of forgetting over a period of weeks or months (Butler & Zeman, 2008). This phenomenon is known as Accelerated Long-term Forgetting (ALF).

1.2 The consolidation process: Models of memory

Earlier models of memory suggest that newly-acquired information, although held temporarily within a short-term memory store, is consolidated into long-term memory within a matter of seconds or minutes (see e.g., Atkinson & Shiffrin, 1968; Baddeley & Warrington, 1970). This is known as the Dual-Trace View. However, more recent consolidation models suggest that although the process of consolidation starts rapidly after presentation of information, full consolidation takes place over a more protracted time period and is unlikely to be complete before at least several hours, days or weeks (e.g., Alvarez & Squire, 1994; Squire & Alvarez, 1995; Squire, Stark, & Clark, 2004). There are two competing theories that propose this: the Standard Model (see e.g., Alvarez & Squire, 1994; Squire & Alvarez, 1995; Squire et al., 2004) and the Multiple Trace Theory (see e.g., Nadel & Moscovitch, 1997; Nadel, Samsonovich, Ryan, & Moscovitch, 2000) (see Extended paper 1.2 for more information on memory consolidation theories). Both theories agree that before memory consolidation is complete, information

remains vulnerable to disturbance. ALF is said to occur due to a disruption to this consolidation process (Hoefeijzers, Dewar, Della Sala, Zeman & Butler, 2013). This fits with the presentation of those with TLE experiencing higher rates of ALF, as the temporal lobes are central to the process of consolidating information from short-term memory to long-term memory (Alvarez & Squire, 1994) (see Extended paper 1.3 for more information on the involvement of the temporal lobes in this process).

1.3 Memory measures

1.3.1 Traditional 'objective' memory measures

Most clinicians still rely on memory measures created on the basis of the Dual-Trace View and, as such, they measure consolidation over a 30-40 minute delay period (asking clients to recall or recognize earlier-presented material at this time period to ascertain forgetting). This is problematic for those who experience ALF, as the delay period is too short to detect their consolidation difficulties. Piazzini, Canevini, Maggiori and Canger (2001) therefore argue that traditional memory tests lack specificity (see Extended paper 1.4.1). This would explain why, when TLE/ epilepsy sufferers complain of ALF-related memory complaints, their subjective accounts do not correlate with results on 'objective' memory measures. For example, correlations between subjective-objective measures have been found to be only moderate (.3-.4) (Brown, Dodrill, Clark & Zych, 1991; O'Shea, Saling, Bladin & Berkovic, 1996) or performance on objective measures is within the normal range (see e.g., Baños et al., 2004; Piazzini et al., 2001; Thompson & Corcoran, 1992).

Furthermore, traditional memory measures may lack ecological validity. This is the ability of the test to predict functional impairment (i.e. an individual's performance in a real-world setting) (Sbordone, & Long, 1996). For example, Hall, Isaac and Harris (2009) suggested that the Wechsler Memory Scale (WMS; a widely used measure clinically) does not measure the memory used for 'everyday' situations. Research suggests memory is multi-modal (see e.g. Annett, McLaughlin Cook & Leslie, 1995; Baddeley, Eysenck, & Anderson, 2009; Bigelow, & Poremba, 2014; Gallace, & Spence, 2009) and so memory tests

artificially separating verbal and non-verbal components may be unrealistic (Helmstaeder, Wietzke, & Lutz, 2009; Wicklund, Johnson, Rademaker, Weitner, & Weintrub, 2006; Zahodne et al., 2011). Recall on multi-modal tasks is perhaps a better indicator of real-life recall (where input generally includes both auditory and visual cues – and sometimes other sensory information). This is particularly relevant for those with ALF-related difficulties, where research suggests that individuals experience consolidation difficulties across modalities (Fitzgerald, Mohamed, Ricci, Thayer, & Miller, 2013) (see Extended paper 1.4.2).

1.3.2 Subjective memory measures

To circumvent these limitations, clinicians might consider using self-report measures as a stand-alone assessment, without the need for traditional objective measures, as self-report measures are generally accepted as having greater ecological validity (Higginson, Arnett, & Voss, 2000). However, this assumes that the person can accurately remember lapses in their memory. This assumption poses problems as people with memory difficulties may in fact struggle to recall lapses in their memory (Sunderland, Harris, & Baddeley, 1983). Helmstaedter, Hauff and Elger (1998) found that subjective memory measures only correlated with objective memory tests in those with intact memories. Furthermore, Hall et al. (2009) reported that subjective memory tests may lack the specificity needed to distinguish memory deficits from difficulties in other cognitive domains. A further difficulty is that people may under- or over- report memory problems according to their motivation, and epilepsy patients' perception of their difficulties, rather than a memory deficit *per se* (Andelman, Zuckerman-Feldhay, Hoffien, Fried, & Neufeld, 2004; Herrman, 1982). This may be understood in terms of deliberate/goal-directed under-reporting (e.g., to avoid confronting difficulties, such as: protecting self and relatives from emotional consequences, or potential functional consequences of 'detection' – employment, independence, relationships, engagement with services) or over-reporting (e.g., to access support). Some subjective biases may also be more state-dependent (e.g., mood congruent [people interpret/recall less favourably when experiencing negative mood]) or trait-dependent (e.g., people differ in response styles, so may tend to endorse more/less extreme categories on measures). Of course, motivated responding is also an issue for objective memory testing. However, the difference

is that it is difficult to 'fake good' on objective testing (whereas one can easily under-report memory difficulties/over-report ability); although it is possible to 'fake bad' on both subjective and objective measures. There are objective measures (e.g., the Test of Memory Malingering [TOMM; Tombaugh, 1996; 1997]) for catching this, and some scales that capture exaggerated self-reporting (e.g., the validity scales of the Millon Clinical Multiaxial Inventory [MCMI; Granacher Jr., 2008; Millon, Millon, Davis, & Grossman, 2009]), but it can be difficult to identify (see Extended paper 1.4.3).

Corbett (2012) accounted for some of the above difficulties by adapting the Memory Functioning Questionnaire (MFQ; Gilewski, & Zelinski, 1988; Gilewski, Zelinski, & Schaie, 1990) for use by carers/relatives, as well as administering it to the patient, and found both patient and carer/relative responses to the MFQ were statistically equivalent overall. However, it is acknowledged that these responses may be somewhat interdependent: patient and carer/relative may have shaped each other's judgement of memory (e.g., developed 'problem' narrative). Corbett (2012) also administered the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) to account for potential impact of mood, with no significant differences being found between clinical and control groups.

1.3.3 ALF-specific memory measures

Consistent with the newer view of consolidation, Blake, Wroe, Breen and McCarthy (2000) compared patients over a period of weeks (as well as the usual immediate and 40 minute delay period) and from this observed that TLE patients showed a faster rate of forgetting. This indicates that using ALF-specific memory measures may be better to detect difference, but these measures are not well-developed or widely available. For example:

(1) Bell (2006) used a story-recall task to assess ALF. Story-recall tasks have limited ecological validity and lack multi-modal input.

(2) Muhlert, Milton, Butler, Kapur and Zeman (2010) attempted to address problems of ecological validity by assessing incidental memory using real-life footage from participants, captured by cameras attached to the participants' heads as they went about their day-to-day activities. This approach, whilst ecologically valid, would be costly to implement across

services and thus is not clinically viable. Furthermore, when using incidental information, it would be very difficult to standardise or interpret performance against broader norms. Memory is difficult to interpret without knowing the individual salience of stimuli and interference/other demands on attention at the time – and it would be unclear what the person attended to/ perceived (as the technique only allowed researchers to see what was in their natural environment). A naturalistic eye-tracking approach may be an advance on this as it would show what the person attended to, however, this approach would still pose a problem with regards to clinical practicality. With instructed learning (i.e., standardised administration with set instructions and format), this can be minimised because we are able to control the conditions and maximise consistency of the learning procedure (although still subject to individual attention/interest).

Testing intervals of existing measures could be adapted, but norms would need to be gathered to assess the reliability and validity of doing so.

1.4 Rationale

Development of a novel measure of ALF was therefore identified as a need within this field. The Accelerated Long-Term Forgetting In Epilepsy (ALFIE) test (Corbett, 2012) aimed to address the key limitations of previous measures: (1) clinical practicality (through administration of the two-week delayed recall and recognition via telephone as well as development of a more cost-effective, standardised testing method) (2) use of stimuli with greater verisimilitude for everyday multi-modal experiences (whilst acknowledging that 'ecological validity' is still limited with respect to whether these stimuli, TV news broadcasts, are individually 'meaningful' or relevant to the everyday functioning difficulties that may prompt referral). By developing norms for long-term (two-week) delayed recall the ALFIE was expected to be specifically useful for measuring ALF and stand out from other memory tests that will likely not have these norms.

In terms of administration, Corbett's (2012) research found that assessing ALF after a 2-week (±2 days) delay can be effectively achieved remotely via the telephone. This method is both practical for the patient and the clinician. Although previous research has also made use of this method (e.g., Bell, 2006), Corbett's

(2012) study addressed key weaknesses in the methodology, such as poor participant matching and test selection. As a result, unlike Bell's (2006) study where there was no indication of ALF or difference in memory between a TLE and control sample, Corbett (2012) found a higher rate of forgetting in TLE patients.

In terms of a multi-modal experience, although it could be argued that previous researchers have provided some of this multi-modal information when, for example, reading stories aloud, the ALFIE presentation is explicitly multi-modal and therefore arguably more engaging. Moreover, delivery is standardised across administrators. Use of news items also provides 'face valid' similarity to stimuli encountered in real life.

Furthermore, as the management of epilepsy is focused not only on minimising patient seizures, but also on reducing any psychological and social difficulties caused by the condition (Browne & Holmes, 2008), National Institute for Health and Clinical Excellence guidelines (2012) suggest that neuropsychological examination should be considered as part of a standard procedure for epilepsy patients. Therefore, it is important that the assessment tools used have the specificity needed, but are also cost-effective and practical for services to use. The ALFIE addresses all of these areas.

Furthermore, as an epilepsy population may have a range of co-morbid psychological needs (see e.g., Hermann et al., 2000), cognitive assessment could help to inform the delivery of interventions for these needs (e.g., making use of recordings and inter-sessional calls/prompts if memory is an issue).

1.5 Aims

The overall aim of the current research was to examine whether the ALFIE is a clinically viable test measure. This was broken down into several components: 1) Examining the validity of the ALFIE (e.g., through concordance of the ALFIE test with a published objective memory measure and assessing test veridicality by ascertaining the relationship between subjective and objective memory), 2) assessing the reliability of the ALFIE measure (e.g., through exploration of the

two form versions and assessment of inter-rater reliability), 3) assessing the acceptability/ practicability of the implemented ALFIE test procedure (with implications for clinical feasibility) (e.g., through assessing attrition rate), and; 4) creating healthy participant standardised norms.

2. Method

This research assumes a positivist epistemological stance, meaning that it was assumed that reality could be observed and measured, and knowledge derived from the study findings that would provide an insight into the objective 'truth' through deliverance of probabilistic evidence (Chalmers, 1999; Giddings, & Grant, 2007; Polit, Beck, & Hungler, 2001, cited in Giddings, & Grant, 2007). Observations were of the decay of newly-learnt material over a two-week period. This research was reviewed and approved by the University of Lincoln Research and Ethics Committee (REC reference: EC07032014).

2.1 Participants

Fifty-three healthy volunteers, aged 18-75 years (based on the age range in Corbett's (2012) study of people with TLE) were recruited (see Extended paper 2.1 for discussion of the sample size and section 2.2 for a discussion of the recruitment process). Nine age categories were recruited to (see Extended paper 2.3 for more information on demographic variables).

2.2 Exclusion criteria

Prospective participants were considered ineligible to participate in the study if they met any of the following exclusion criteria: 1) aged <18 or ≥75 years, 2) non-English speakers (as the neuropsychological measures being used had not been validated on this population), 3) uncorrected vision or hearing loss (due to the modality of material presentation), 4) any diagnosis of epilepsy, progressive neurological disorder (e.g., multiple sclerosis, dementia, gliomas of grade two or above), cerebrovascular infarction, traumatic brain injury, encephalitis, or mood disorder, 5) any past or present neurosurgery or memory rehabilitation, 6) currently on psychotropic medication, and 7) those who lacked ability to give informed consent at any stage of the research. All inclusion/ exclusion criteria were determined by self-report.

2.3 Objective memory assessment

Participants completed a battery of objective memory measures (for justification of test selection and discussion of the psychometric properties of published tests, see Extended paper 2.4).

2.3.1 BMIPB memory measures

The Brain Injury Rehabilitation Trust (BIRT) Memory and Information Processing Battery (BMIPB; Coughlan, Oddy, & Crawford, 2007) was used to assess participants' recall and recognition memory for verbally presented information. The advantage of the BMIPB over other objective memory measures (e.g., WMS-III, Wechsler, 1997a; previously used by Bell, 2006) was that it has several 're-test versions,' similar to the ALFIE. Therefore, if the ALFIE was not found to be a valid and reliable test measure then the normative data collated in this study might be used to create an extension of the BMIPB measure in order to capture ALF-type difficulties. Alternate test versions are important for clinical samples who may be re-tested at a later date, as this minimises the chance of content-specific practice effects that may occur due to test familiarity, something that has been highlighted as problematic in previous research designs (Jansari, Davis, McGibbon, Firminger, & Kapur, 2010). Both Form 1 (V1) and Form 2 (V2) of the BMIPB have been shown to be equally difficult (Coughlan et al., 2007). In the current study, counterbalancing was used to ensure an equal number of participants completed each form. Where participants were recruited who knew each other, alternate form versions were used wherever possible to prevent joint rehearsal between time-points. The following two sub-tests of the BMIPB were used:

2.3.1.1 BMIPB Story recall

Participants were read a short story (e.g., about two lions escaping from a zoo) and then free recall was assessed immediately (T1), after a 40-minute delay (T2) and after a two-week delay (T3). The two-week delay is not part of the formal test procedure for the BMIPB, but was created specifically for this research to enable a comparison to the two-week delay of the ALFIE. Scoring was conducted in line with the BMIPB manual (maximum score possible at each time-point = 60, higher scores indicating better memory recall).

2.3.1.2 BMIPB Word List recall and recognition

Participants were read a list of 15 words (e.g., Village, Dust, Frog) and asked to freely recall as many of the words as possible, in any order, immediately after presentation. This learning procedure was carried out five times, regardless of the level of learning achieved. Scores for this part of the test were treated as part of the learning process and were not analysed. Free recall of a 15-word distracter list was then assessed once. Without stimulus re-presentation, the participant was then asked to recall the original list immediately (T1), after a 40-minute delay (T2) and after a two-week delay (T3) (maximum score possible at each time-point = 15, higher scores suggest better memory recall).

Following both T2 and T3 delayed recalls the participant was also assessed for word and list recognition. They were presented with a list of 30 word-pairs comprised of an original word (from the first list that they heard five times and the distracter list that they heard once) and a novel distracter word (e.g., Bank-Money). They were asked to correctly identify the word they had previously heard and determine which list it had been in. Participants were asked to guess if they were unsure. Distracter words used at T2 and T3 were different, to avoid any confusion over word-recognition of distracter words. Distracter words at two-week delay were comprised of words from the alternate BMIPB form (either V1 or V2 accordingly) as constructed by Corbett (2012).

2.3.2 ALFIE Story recall and recognition

The ALFIE design was similar to that of the BMIPB, with components of both recall and recognition memory assessed over different time-points. However, it was expected that the ALFIE would have greater ecological validity than the BMIPB due to the stimuli being sourced from real televised BBC news broadcasts (see Extended paper 2.5 for permissions regarding this material). As with the BMIPB, two versions were used (V1 and V2), with counterbalancing to ensure an equal number of participants completed each version.

Participants were shown a short audio-visual news clip (2-3 minutes in length) on a laptop computer and asked to recall the story immediately (T1), after a 40-minute delay (T2) and after a two-week delay (T3). V1 tells the story of a man

who was fined for brushing his dog in the park and V2 is a story about a woman who cycled and kayaked around the globe. Both stories had the same number of information units within them to ensure equal task complexity. An information unit was a piece of information that could be tapped through a recognition memory question. For example, 'The *man stood* by the *door*' contains three information units (italicised) whereas 'The *man stood* by the *door searching* for his *keys*' contains five information units. Based on the number of information units Corbett (2012) developed a 28-item score sheet to assess free recall (maximum score possible at each time-point = 28, higher scores indicating better memory recall) (see Appendix C for ALFIE story score sheets and Appendix D for ALFIE story scoring guidance).

Following both delayed recalls (T2 and T3) participants were asked 'yes/no' questions to assess recognition memory. Fourteen questions were asked about details regarding the narrative content (e.g., Was the man's name Roy Wyer?) (maximum score possible at each time-point = 14, higher scores indicating better recognition memory) and 14 questions were asked about visual features present within the news clips (e.g., Were the presenters sitting on a blue sofa?) (maximum score possible at each time-point = 14, higher scores suggest better recognition memory). Each recognition question targeted one information unit from the story (see Appendix E for ALFIE Recognition score sheets).

2.4 Subjective memory assessment

The Everyday Memory Questionnaire-28 (EMQ-28; Sunderland, Harris, & Gleave, 1984) was used to assess participants' memory via self-report. Participants were asked to rate 28 statements related to everyday memory failure (e.g., Telling someone a story or joke that you have told them once already) according to how they felt their memory had been over the last three months. An additional item (Q29: "Forgetting something you have said or done two weeks previous") was added to provide a subjective rating regarding the participant's experience of ALF-type difficulties. Items were rated by the participant using a 9-point Likert-scale, ranging from 0: "Not at all in the last six months" to 8: "More than once a day." Higher overall scores on testing indicated poorer subjective memory performance (maximum score possible = 232). The EMQ-28 is quick and

simple to administer due to its questionnaire format and has been shown to have good response and construct validity, and good internal consistency (Cronbach's α=.85-.91) (Cornish, 2000; Efklides et al., 2002; Royle & Lincoln, 2008; Sunderland, Harris, & Gleave, 1984).

2.5 Demographic measures

Participants' date of birth, gender and level of educational attainment were recorded. A test to ascertain Predicted Full-Scale IQ (PFSIQ) was also conducted.

2.5.1 WTAR

The Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) was used to ascertain PFSIQ. Participants were given a sheet of 50 irregularly spelt/ irregularly pronounced words and asked to read them out loud (e.g., 'knead,' 'paradigm'). Participants were encouraged to guess if unsure (maximum score possible = 50, higher scores indicated higher PFSIQs). Raw scores were converted to standard scores (PFSIQs) using the WTAR manual. The WTAR has good psychometric properties, with high levels of reliability and validity (see e.g., Wechsler, 1997b, 2001).

2.6 Data collection procedure

The study procedure was the same for all participants, with neuropsychological assessment taking place over two sessions, as detailed below.

2.6.1 Stage 1: Initial assessment

The initial assessment was conducted face-to-face with the primary author. All participants completed the BMIPB story recall, the BMIPB word list recall and the ALFIE story recall tasks immediately (T1) and after a 40-minute delay (T2). Learning at T1 was assessed via free recall only. At T2 participants completed both free recall and recognition tasks. Between T1 and T2 participants completed the EMQ-28 and the WTAR. Participants were not required to do anything between Stage 1 and Stage 2 (two-week follow up) (see Extended paper 2.2.5 for further discussion on participant instructions for between Stage 1 and Stage 2 of the research).

2.6.2 Stage 2: Assessment after extended delay period

Two weeks ±2 days after the initial assessment, the participant was contacted by telephone. Memory of the BMIPB story, the BMIPB word list and the ALFIE story was assessed via free recall and recognition tasks (T3).

2.7 Approach to statistical analyses

All analyses were conducted using SPSS Version 22 statistics software. Parametric statistics were employed for the BMIPB story and the ALFIE recognition tasks as data met the assumptions for this family of tests. All other tasks were analysed using non-parametric statistics or through bootstrapping of scores. Single case representatives were excluded for mixed ANOVAs (see Extended paper 2.6 for an expansion on the approach to statistical analysis).

3. Results

All 53 participants completed assessment at T1 and T2. Two participants did not complete the second assessment session and were therefore not included in any T3 analysis. There was an uneven sample size, with one more participant completing V1 of the objective memory tests (n = 27) than the V2 equivalent (n = 26). With the exclusion of the participant outside the normative IQ range (see below) the number of participants who completed V2 decreased by one (n = 25).

Preliminary analysis revealed no effect of age on memory performance, and so examination of demographic variables was conducted using the total sample, as opposed to splitting by pre-defined age groups. Overall the sample were representative of the normal population: Approximately half of the participants (n = 24) performed between the 25^{th}-75^{th} percentile range of the established objective memory measure (BMIPB) at immediate recall, they were of an average education in comparison to the literature (see e.g., The Organisation for Economic Co-operation and Development, 2014) and of an average Predicted Full-Scale IQ (PFSIQ), except for one participant who was removed from the dataset and excluded from analysis for statistical and conceptual reasons, as their PFSIQ came outside of the normative range (an IQ of <70) (See Table 6; see Extended paper 3.1 for discussion on exclusion of participants on the basis of PFSIQ).

Table 6. Demographic and background information

Age category	Gender (male:female)	Age (mean in years)	Education (mean in years)	WTAR (PFSIQ)
18.0-19.11	1:7	18.63 (0.52)	14.50 (0.93)	106.13 (8.66)
20.0-24.11	3:2	21.8 (1.30)	17.40 (2.19)	114.40 (8.17)
25.0-29.11	5:4	26.44 (1.33)	18.50 (1.94)	107.56 (9.99)
30.0-34.11	5:0	32.0 (1.58)	17.20 (2.95)	114.00 (7.62)
35.0-44.11	2:4	38.67 (2.42)	17.33 (5.72)	97.83 (9.04)
45.0-54.11	1:3	52.50 (1.29)	11.75 (0.96)	95.50 (14.85)
55.0-64.11	3:5	58.38 (4.07)	12.75 (2.66)	105.50 (10.71)
65.0-69.11	3:2	66.50 (1.00)	14.25 (3.86)	94.00 (26.88)
70.0-74.11	2:1	71.67 (1.16)	11.00 (1.00)	111.00 (14.18)
Whole sample (18.0-74.11)	24:28	39.35 (18.04)	15.36 (3.64)	105.85 (10.99)

Standard deviations shown in parentheses; WTAR = Wechsler Test of Adult Reading (Wechsler, 2001); PFSIQ = Predicted Full-Scale IQ.

3.1 Reliability

The following section will assess the reliability in two ways: 1) examination of parallel forms reliability, through exploration of memory performance on V1 and V2 (and any impact of demographic variables on this) (section 3.1) and 2) examination of inter-rater reliability, by analyse of level of agreement between independent markers (section 3.2).

3.1.1 Parallel forms reliability

3.1.1.1 Memory performance across test versions

Mann-Whitney U tests and independent t-tests (with test version [V1 vs. v2] as the independent variable) indicated a significant difference in participant performance on all objective memory recall test versions at T1, which remained after Holm's sequential Bonferroni method for multiple comparisons was applied (.05/3 or 3 comparisons = significance *p* value of ≤.05, .025 or .017 respectively; ALFIE Story: $U = 44$, $p = .01$; BMIPB List recall: $U = 173$, $p = .01$; BMIPB Story: $t(48) = -2.17$, $p = .04$).

For recognition tasks (tested at T2) an independent t-test found a significant difference in test version performance for the ALFIE narrative recognition task ($t(50) = 6.66$, $p = .01$). Mann-Whitney U tests and an independent t-test found no significant differences in test version performance for all other recognition tasks (ALFIE visual recognition: $t(50) = .22$, $p = .82$; BMIPB List word and list recognition: $U = 282.00 – 276.50$, $p \geq .25$).

As significant differences were found between V1 and V2 test performance, Chi-square tests were employed to examine whether those who completed V1 and V2 were equivalent in terms of their demographic variables (age, gender, years of education). No significant associations were found between demographic variables and test version administered (X^2 (1 – 8) = .08 – 10.78, $p \geq .22$). Propensity score matching was then conducted to enable matching of observed variables. Propensity score matching reduced differences across observed covariates, with no covariates exhibiting a large imbalance between test versions (where 'large' was considered to be $d > \pm .25$). V1 participants were then paired with V2 participants (according to nearest 'neighbour' on propensity score) and data treated as dependent. Paired analyses produced similar results to chi-square tests above, although the BMIPB Story task did not reach significance after matching ($t(22) = 1.90$, $p = .07$). This indicated that the parallel forms reliability of ALFIE test versions was poor and they could not be taken as parallel forms. All subsequent analysis of ALFIE raw data and BMIPB List raw data therefore split the data into two groups (V1/ V2).

3.1.1.2 Memory performance and demographic variables

The analyses in the section will examine 1) whether memory performance was influenced by individual differences in age, gender, education, or IQ to enable creation of the standardisation sample, and 2) whether tests were sensitive to (expected) changes in memory performance over time.

3.1.1.2.1 BMIPB Story recall

Mixed ANOVAs were conducted to examine the effect of gender, age and years of education on BMIPB Story scores at T1, T2 and T3. Mauchly's test indicated that the assumption of sphericity was violated ($X^2(2) = 9.64$, $p = .01$), therefore

degrees of freedom were corrected using Greenhouse-Geisser estimates (ε = .70). No significant effect of any demographic variables was found at any time point (F(2 – 16) = .04 – .47, p ≥.87). Pearson's correlation coefficient was used to look for an association between PFSIQ (As determined by the WTAR) and BMIPB Story scores at T1, T2 and T3. No significant relationship was found at any time point (r(52) = -.09 – -.01, p ≥.52).

Paired samples t-tests were used to test for differences in participant's scores between assessment time points (T1-T2, T2-T3 and T1-T3). A significant difference was found between scores all time points (T1-T2: t(51) = 4.26, p = .01; T2-T3: t(49) = 8.23, p = .01; T1-T3: t(49) = 9.16, p = .01). This indicated decrements in memory performance between successive time-points (immediate, 40-minute delay and two-week delay) (see Table 7).

3.1.1.2.2 BMIPB Word List recall

Mixed ANOVAs were conducted to examine the effect of gender, age and years of education on BMIPB V1 and V2 List recall scores at T1, T2 and T3[1]. A mixed ANOVA indicated a significant effect of age on BMIPB V1 List recall scores (F(12) = 2.75, p = .04). However, no significant effect of age on BMIPB V1 List recall scores was found after Bonferroni post-hoc testing (p ≥.54). Results regarding the effect of age were therefore taken as non-significant due to Bonferroni being a more conservative test of difference (as it attempts to control the overall alpha level). It is possible that the significant effect of age found by the mixed ANOVA was a Type I error resulting from a small sample size. No significant effect was found for all other demographic variables for V1 (F(2 – 6) = .97 – 1.08, p ≥.40) and for V2 (F(2 – 10) = .29 – 1.17, p ≥.40).

Bootstrapped Pearson's correlation coefficients were conducted to examine the association of PFSIQ with BMIPB V1 and V2 List recall scores at T1, T2 and T3.

[1] Kruskal-Wallis tests were also conducted on the data and the outcome was the same as reported above. Therefore, although we acknowledge that some assumptions for parametric testing were not met, the results of mixed ANOVAs have been reported due to having more precision and power. This method was also applied to the BMIPB List recognition tasks and the ALFIE Story recall (See Extended section 2.6.3 for further information).

No significant relationships were noted (V1: $r(25) = -.27 – -.03$, 95% CI [-.67 – .39], $p \geq .20$; V2: $r(25) = -.24 – .26$, 95% CI [-.58 – .54], $p \geq .22$).

Wilcoxon signed-ranks tests were used to look for difference in participant scores between assessment time points (i.e. between: T1-T2, T2-T3 and T1-T3). For participants who completed V1 and V2 there were significant differences between scores T2-T3 (V1: $z = -4.38$, $p = .01$; V2: $z = -4.31$, $p = .01$) and T1-T3 (V1: $z = -4.02$, $p = .01$; V2: $z = -4.29$, $p = .01$). This indicated decrements in memory performance between immediate recall/ 40-minute delay and the two-week delay. No significant differences were noted between scores on either V1 or V2 between T1 and T2 ($z = -.50 – -.86$, $p \geq .39$). This indicated no significant decrement in memory performance between immediate recall and the 40-minute delay (See Table 7).

3.1.1.2.3 BMIPB Word List recognition tasks

Mixed ANOVAs indicated a significant effect of all demographic variables on BMIPB V1 List list-recognition scores (Gender: $F(1) = 81.38$, $p = .01$; Age: $F(6) = 53.94$, $p = .01$; Years of education: $F(3) = 8.32$, $p = .01$). However, Bonferroni post-hoc testing did not reveal any significant differences for age ($p \geq .30$) and years of education ($p > .99$) and therefore results regarding the effect of these demographics were taken as non-significant. For BMIPB V1 List word-recognition scores mixed ANOVAs indicated a significant effect of gender ($F(1) = 7.64$, $p = .03$). There was no significant effect of age or years of education ($F(3\text{-}6) = 2.24 – 3.66$, $p \geq .06$). For BMIPB V2 List list-recognition and BMIPB V2 List word-recognition no effect of demographic variables was found ($F(1\text{-}5) = .03 – 3.7$, $p \geq .09$).

Bootstrapped Pearson's correlation coefficients were used to examine the association between PFSIQ and BMIPB V1 and V2 List list-recognition and List word-recognition scores. No significant relationships were noted (V1: $r(25) = -.16 – .21$, 95% CI [-.71 – .64], $p \geq .30$; V2: $r(25) = -.07 – .32$, 95% CI [-.45 – .64], $p \geq .12$).

Wilcoxon signed-ranks tests revealed significant differences in participant scores between T2 and T3 for the BMIPB List list-recognition task (V1 males: z = -2.67, p = .01; V1 females: z = -3.05, p =.01; V2: z = -3.86, p = .01). However, no significant differences were noted in participant scores between T2 and T3 for the BMIPB List word-recognition task (z = -1.46 – -.18, p ≥.15). This indicated decrements in memory performance between successive time-points (40-minute delay and two-week delay) for list-recognition tasks, but not word-recognition tasks (See Table 7).

Table 7. Mean group scores (and standard deviations) for the BMIPB objective memory tests

BMIPB memory tests			
Story recall	V1 or V2		
Immediate (T1)	29.02 (10.04)		
40-minutes (T2)	27.08 (10.23)		
2-weeks (T3)	19.66 (9.01)		
Word List recall	V1		V2
Immediate (T1) recall	12.63 (2.24)		10.64 (3.16)
40-minutes (T2) recall	12.48 (2.23)		10.48 (3.42)
2-weeks (T3) recall	7.04 (3.54)		5.16 (3.75)
Word List recognition	V1 males	V1 females	V2
List-recognition 40-minutes (T2)	27.67 (2.10)	28.73 (1.98)	27.00 (3.87)
List-recognition 2-weeks (T3)	24.09 (4.00)	23.79 (5.70)	23.28 (3.87)
Word-recognition 40-minutes (T2)	26.92 (2.54)	27.40 (1.72)	26.72 (2.07)
Word-recognition 2-weeks (T3)	26.64 (2.34)	26.21 (2.81)	26.12 (2.57)

BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); V1 = Test Version 1; V2 = Test Version 2.

3.1.1.2.4 ALFIE Story recall

Mixed ANOVAs were used to examine the effect of demographics variables on test version performance. A significant effect of gender on ALFIE Story V1 scores was found (F(2) = 8.22, p = .01). All other demographic factors were found to be

non-significant at the .05 level (F(6 – 12) = 1.06 – 2.17, p ≥.10). For the ALFIE Story V2 there was a significant effect of age (F(10) = 4.54, p = .01) and years of education (F(4) = 5.09, p = .02)[2] on ALFIE Story V2 scores. The effect of gender was non-significant at the .05 level (F(2) = .60, p = .57). However, Bonferroni post-hoc testing did not reveal any significant differences between multiple comparisons (p ≥.61) and therefore initially 'significant' differences did not survive correction for multiple comparisons.

Bootstrapped Pearson's correlation coefficients were conducted to examine the relationship of PFSIQ and ALFIE Story V1 and V2 scores. No significant relationships were noted (V1: r(25) = .24 – .25, 95% CI [-.25 – .58], p ≥.26; V2: r(25) = .03 – .14, 95% CI [-.50 – -.35], p ≥.52).

Scores for V1 males, V1 females and V2 participants were then analysed using Wilcoxon signed-ranks tests to examine any difference in ALFIE Story scores over time (between T1-T2, T2-T3 and T1-T3). All participant scores showed significant differences between T2-T3 (V1 males: z = -2.71, p = .01; V1 females: z = -3.14, p = .01; V2: z = -4.03, p = .01) and T1-T3 (V1 males: z = -2.82, p = .01; V1 females: z = -2.89, p = .01; V2: z = -4.29, p = .01). No significant differences were found between T1-T2 (z = -1.93 – -1.18, p ≥.054), although the ALFIE Story V2 did approach significance (z = -.20, p = .054). This indicated a decline in memory scores between immediate recall/ 40 minute delay and the two-week delay, but not significant decrement in memory performance between immediate recall and the 40-minute delay (see Table 8).

Association between performances at different time-points of the ALFIE Story was examined using bootstrapped Pearson's correlation coefficients. Significant relationships at all time points were found for both V1 males (T1-T2: r(11) = .87, 95% CI [.55 – .98], p = .01; T2-T3: r(11) = .81, 95% CI [.08 – .97], p = .01; r(11) = .72, 95% CI [-.04 – .93], p = .01) and V2 participant scores (T1-T2: r(25) = .79, 95% CI [.66 – .88], p = .01; T2-T3: r(25) = .52, 95% CI [.23 – .75], p = .01; T1-T3:

[2] An alpha level of .025 was utilised for ALFIE Story V2 T3 years of education ANOVA analysis (See Extended section 2.6.2.1.4).

r(25) = .65, 95% CI [.46 – .82], *p* = .01). For V1 females, there was only a significant association between scores at time points T1-T2 (*r*(14) = .75, 95% CI [.33 – .93], *p* = .01), but not between T2-T3 or T1-T3 (*r*(14) = -.12 – -.22, 95% CI [-.62 – .47], *p* ≥.46).

3.1.1.2.5 ALFIE Recognition tasks

Mixed ANOVAs revealed no significant effect of gender, age or years of education on ALFIE V1 or V2 narrative recognition scores at T2 or T3 (*F*(1-6) = .68 – 4.70, *p* ≥.07). For ALFIE V1 visual recognition scores at T2 and T3 there was a significant effect of gender (*F*(1) = 5.86, *p* = .05), but no significant effect of age or years of education (*F*(3 – 6) = .33 – 3.5, *p* ≥.06). To make participants scores comparable, recognition tasks have all been subsequently split into gender groups for V1. For ALFIE V2 visual recognition scores at T2 and T3 there was a significant effect of years of education (*F*(2) = 5.88, *p* = .05). However, this significant effect became non-significant when Bonferroni post-hoc testing was applied (*p* ≥.70) and therefore results that showed initially 'significant' differences did not survive correction for multiple comparisons. There was no significant effect of gender or age (*F*(1 – 5) = 1.44 – 3.71, *p* ≥.09) on ALFIE V2 visual recognition scores.

The effect of PFSIQ on ALFIE V1 and V2 recognition tasks at all time points was examined through bootstrapped Pearson's correlation coefficients. There was no significant association for either the ALFIE V1/V2 narrative recognition and visual recognition (V1: *r*(25) = -.01 – .36, 95% CI [-.31 – .64], *p* ≥.08; V2: *r*(25) = -.12 – .21, 95% CI [-.58 – .60], *p* ≥.32).

Paired samples t-tests found a significant difference in V1 females participant scores between T2 and T3 for the ALFIE narrative recognition (*t*(13) = 2.86, *p* = .01) (see Table 8). No other significant differences were noted for ALFIE recognition tasks (for both V1 males, V1 females and V2 participants) (Narrative recognition: *t*(10 – 24) = .01 – .39, *p* ≥.70; Visual recognition: *t*(10 – 24) = 1.25 – 1.88, *p* ≥.09). This indicated decrements in memory performance between successive time-points (40-minute delay and two-week delay) for females on the ALFIE V1 narrative recognition task (see Table 8).

Pearson's correlation coefficients indicated that narrative recognition tasks showed an association between performance at different time points for V1 females ($r(14) = .76$, $p = .01$) and V2 participants ($r(25) = .56$, $p = .01$), but not for V1 males ($r(11) = .43$, $p = .19$). Visual recognition tasks only showed an association between performance at different time points for V2 participants ($r(25) = .54$, $p = .01$), for all other test versions no significant relationship was found ($r(11\text{-}14) = .40 – .45$, $p \geq .16$).

Table 8. Mean group scores (and standard deviations) for the ALFIE objective memory tests

ALFIE memory tests	V1 males	V1 females	V2
Recall			
Immediate (T1)	14.33 (3.34)	14.47 (3.70)	7.88 (2.44)
40-minutes (T2)	13.50 (3.66)	15.20 (3.26)	7.24 (2.17)
2-weeks (T3)	10.18 (3.40)	9.57 (2.10)	4.12 (2.56)
Recognition			
Narrative 40-minutes (T2)	12.50 (1.09)	12.80 (.94)	10.36 (1.47)
Narrative 2-weeks (T3)	12.64 (1.03)	12.14 (1.29)	10.24 (1.76)
Visual 40-minutes (T2)	10.42 (0.90)	9.53 (1.30)	9.84 (1.55)
Visual 2-weeks (T3)	9.82 (1.17)	9.00 (1.71)	9.48 (1.45)

ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); V1 = Test Version 1; V2 = Test Version 2.

3.1.2 Inter-rater reliability

A sub-section of ALFIE participant forms ($n = 5$; 10%) were scored independently by a secondary marker to ascertain inter-rater reliability. Out of 700 items (available marks across the five participants) both markers were in agreement for 697 items (i.e., had both marked consistently, as correct or incorrect), meaning percent agreement of 99.57%. This equated to a kappa of .99.

3.2 Validity

Validity was analysed by firstly creating percentile norms for the ALFIE and the extended sections of the BMIPB (section 3.2.1), which allowed assessment of convergent validity of the ALFIE with the BMIPB through percentiles matching

(section 3.2.2). Then, concordance of the ALFIE memory test scores with results on the subjective memory measure were examined (section 3.3.3).

3.2.1 Creation of percentile norms

3.2.1.1 ALFIE percentiles

ALFIE V1 and V2 raw scores were standardised by separately converting them into comparable percentiles (see Extended paper 3.2.1 for discussion on choice of transforming to percentiles). As there was a significant effect of gender on ALFIE V1, participant data was split into male and female before the transformation of raw scores. For the ALFIE V2, participant data was standardised without any splitting by demographic variables (See Appendix F for ALFIE percentile tables).

T2 and T3 scores were also expressed as percentages of the time point before them to examine how much participants retained, based on their initial learning (i.e. T2 score expressed as a percentage of T1 score, T3 score expressed as a percentage of T2 score). T3 score was also expressed as a percentage of T1 score to examine retention across the whole testing period. These were then transformed into percentiles to enable standardisation.

Landmark percentiles were chosen in line with current objective memory measures (e.g., the BMIPB). Norms therefore consisted of tables presenting raw scores that corresponded to the 2^{nd}, 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles (see Extended paper 3.2.2 for percentile definition).

3.2.1.2 BMIPB percentiles

BMIPB percentiles were needed for some T2 and all T3 tasks, as current BMIPB procedures did not measure memory at these time points. Percentiles were calculated using the same method as for the ALFIE. For the BMIPB Story percentiles could be created together for V1 and V2. For all other BMIPB sub-tests, raw scores had to be separated before being expressed as percentiles (BMIPB List recall into V1 and V2, BMIPB recognition tasks into V1 males, V1 females and V2). (See Appendix G for BMIPB extended percentile tables.)

3.2.2 BMIPB and ALFIE percentiles matching

Bootstrapped Pearson's correlation coefficients were used to examine the relationship between percentile ranks obtained from test scores on the BMIPB and the ALFIE at T1, T2 and T3.

3.2.2.1 Free recall tasks

Bootstrapped Pearson's correlations indicated a significant positive relationship between the BMIPB Story and the ALFIE Story percentiles at all time-points (T1: $r(49) = .38$, 95% CI [.14 – .58], $p = .01$; T2: $r(49) = .44$, 95% CI [.20 – .64], $p = .01$; T3: $r(49) = .16$, 95% CI [-.05 – .40], $p = .01$). Bootstrapped Pearson's correlations also found a significant positive relationship between the BMIPB List recall and the ALFIE Story percentiles at all time-points (T1: $r(49) = .55$, 95% CI [.33 – .72], $p = .01$; T2: $r(49) = .32$, 95% CI [.08 – .53], $p = .02$; T3: $r(49) = .32$, 95% CI [.03 – .54], $p = .03$).

3.2.2.2 Recognition tasks

Bootstrapped Pearson's correlations found no significant association between any of the BMIPB Recognition tasks (List: List and Word) and any of the ALFIE Recognition tasks (Narrative and Visual) at any time point (T2: $r(49) = -.11 – .16$, 95% CI [-.36 – .44], $p \geq .27$; T3: $r(49) = -.10$, 95% CI [-.40 – .34], $p \geq .49$).

3.2.3 Subjectively measured memory

All participants rated their own memory via the EMQ-28 with a higher total score indicating greater perceived difficulties with participant's memory. Association between EMQ-28 scores and objective memory assessment scores could therefore be examined. Bootstrapped Pearson's correlation coefficients were conducted to examine the relationship between subjective memory scores (EMQ-28) and objective memory "difference" scores (i.e. the difference between the score obtained by participants at T1-T2, T2-T3 and T1-T3). Significant associations were found between the EMQ scores and the BMIPB List Recall T2-T3 and T1-T3 difference scores (T2-T3: $r(25) = -.40$, 95% CI [-.66 – -.05], $p = .05$; T1-T3: $r(25) = -.43$, 95% CI [-.69 – -.10], $p = .03$). For the ALFIE, significant relationships were found between the EMQ score and the ALFIE V1 Story T2-T3 difference score ($r(11) = -.79$, 95% CI [-.93 – -.58], $p = .01$) for males, and also

between the EMQ-28 score and the ALFIE V1 visual recognition task T2-T3 difference score ($r(13) = -.57$, 95% CI [-.86 – -.27], $p = .04$) for females. No other significant relationships were noted for correlations with BMIPB or ALFIE sub-tests. However, due to small sample sizes effect size *r* was used to determine effect size (irrespective of significance). When effect size was examined all of the ALFIE sub-tests showed at least a 'small' positive association with performance (Cohen, 1992), with 60% of the ALFIE sub-test difference scores showing greater than or equal to a "moderate" effect size, compared to 20% of the BMIPB sub-test difference scores, with an equal percentage of BMIPB sub-test difference scores showing 'trivial' effect sizes (see Table 9).

In addition to the overall raw score of the EMQ-28, bootstrapped Pearson's correlation coefficients were similarly conducted between objective memory measure difference scores (at T2-T3 and T1-T3) and the rating given to a specific question related to ALF-type memory difficulties (EMQ-28 Q29). No significant relationships were found between the novel ALF-specific item and any of the ALFIE or BMIPB sub-test difference scores. Effect size was also examined using effect size *r* (see Table 10). 75% of the ALFIE sub-tests showed at least a 'small' positive association with performance on Q29 (Cohen, 1992). A third (33.33%) of both the ALFIE and the BMIPB sub-test difference scores showed greater than or equal to a moderate effect size with the novel ALF-specific item. These larger effect sizes were only seen when examining T2-T3 difference scores.

Table 9. Effect size *r* for bootstrapped Pearson's correlations with EMQ-28

Memory test correlated with the EMQ-28		Effect size *r*		
		V1 males	V1 females	V2
ALFIE	Recall			
	T1-T2	**	*	*
	T2-T3	***	*	**
	T1-T3	***	**	**
	Recognition			
	Narrative T2-T3	*	**	*
	Visual T2-T3	*	***	**
BMIPB	Story		V1 or V2	
	T1-T2			
	T2-T3		*	
	T1-T3		*	
	List recall		V1	V2
	T1-T2		*	
	T2-T3		**	*
	T1-T3		**	*
	List recognition	V1 males	V1 females	V2
	List T2-T3	*	***	*
	Word T2-T3	*	*	

*EMQ-28 = Everyday Memory Questionnaire (Sunderland, Harris, & Gleave, 1984); ALFIE = Accelerated Long-term Forgetting in Epilepsy test (Corbett, 2012); BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; No asterisk = trivial (<.1); * = small (≥.1); ** = moderate (≥.3); *** = large (≥.5).*

Table 10. Effect size *r* for bootstrapped Pearson's correlations with the novel ALF-specific item (EMQ-28 Q29)

Memory test correlated with the novel ALF-specific item		Effect size *r*		
		V1 males	V1 females	V2
ALFIE	Recall			
	T2-T3	**	*	
	T1-T3	*	*	*
	Recognition			
	Narrative T2-T3		**	
	Visual T2-T3	**	***	*
BMIPB	Story		V1 or V2	
	T2-T3		*	
	T1-T3		*	
	List recall		V1	V2
	T2-T3			**
	T1-T3		*	*
	List recognition	V1 males	V1 females	V2
	List T2-T3	***	**	
	Word T2-T3	*	**	*

*ALF = Accelerated Long-term Forgetting; ALFIE = Accelerated Long-term Forgetting in Epilepsy test (Corbett, 2012); BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; No asterisk = trivial (<.1); * = small (≥.1); ** = moderate (≥.3); *** = large (≥.5).*

3.3 Comparison with a clinical population

Corbett (2012) tested the ALFIE against the BMIPB using both clinical and control samples. Her clinical data has therefore been analysed to determine whether a TLE population would be categorised as experiencing forgetting outside of the normative range when compared to the ALFIE's current standardisation sample. Difference scores (i.e. the rate of forgetting between T1-T2 and T2-T3) were examined by calculating percent retained and then using the ALFIE percentile norms tables in Appendix F (see Table 11).

Concordance between the BMIPB and the ALFIE 'categorisation' (i.e. as experiencing a normal or abnormal rate of forgetting) for the TLE sub-group was also explored through bootstrapped Pearson's correlation coefficients (using established norms for the BMIPB and current norms for T2-T3 BMIPB, and using current norms for the ALFIE).

For free recall tasks, bootstrapped Pearson's correlation coefficients indicated no significant relationship between the BMIPB Story and the ALFIE Story percentiles at T1 ($r(14) = .28$, 95% CI [-.42 – .84], $p = .34$), but a significant positive relationship at T2 ($r(14) = .74$, 95% CI [.35 – .91], $p = .01$) and T3 ($r(14) = .66$, 95% CI [.12 – .90], $p = .01$). Bootstrapped Pearson's correlation coefficients also found a significant positive relationship between the BMIPB Word List recall and the ALFIE Story percentiles at all time-points (T1: $r(14) = .65$, 95% CI [-.07 – .89], $p = .01$; T2: $r(14) = .72$, 95% CI [.12 – .89], $p = .01$; T3: $r(14) = .53$, 95% CI [-.03 – .87], $p = .05$).

For recognition tasks, bootstrapped Pearson's correlation coefficients found no significant association between the BMIPB List: Word Recognition task and any of the ALFIE Recognition tasks (Narrative and Visual) at any time-point (T2: $r(14) = .29 – .31$, 95% CI [-.25 – .89], $p \geq .28$; T3: $r(14) = .06 – .25$, 95% CI [-.61 – .65], $p \geq .39$).

Table 11. Difference scores depicting TLE participants exhibiting forgetting outside of a normative percentile range on the ALFIE

	Test Version	ALFIE Story		ALFIE narrative recognition	ALFIE visual recognition	BMIPB Story		BMIPB List Recall		BMIPB List Recognition Word
		T1-T2	T2-T3	T2-T3	T2-T3	T1-T2	T2-T3	T1-T2	T2-T3	T2-T3
TLE1	V1f			*			*			**
TLE2	V2	*			*					**
TLE3	V1m	*	**		**	*		*	**	*
TLE4	V2			*		*			**	
TLE5	V1f	*		**		*	*	**		**
TLE6	V2							*	**	
TLE7	V2	*		*						
TLE8	V1m	**	*		**					
TLE9	V1m				*	**	*	*		*
TLE10	V2	*		*	*	*		*	**	
TLE11	V2	*	**		*		*	*	**	*
TLE12	V1f	*		**		*	**			**
TLE13	V1f	*	**		*		*	*		*
TLE14	V2				**		*	*		*

*TLE = Temporal Lobe Epilepsy; TLE data from Corbett's (2012) research; ALFIE = Accelerate Long-term Forgetting In Epilepsy test (Corbett, 2012); V1f = Test Version 1 female participants; V1m = Test Version 1 male participants; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; * Indicates memory performance <25th percentile; ** Indicates memory performance >75th percentile.*

4. Discussion

In this study we examined the psychometric properties of a novel measure of ALF, the ALFIE, as compared to an established clinically-used objective memory assessment and a subjective memory measure. This allowed exploration of the validity and reliability of this novel measure. Implications for clinical practicality were also examined. Ecological validity and practicality within a clinical context were both highlighted as difficulties that previous ALF measures have struggled with.

We found provisional evidence for reliability of the ALFIE (in terms of high inter-rater reliability and possible parallel forms reliability once adjusted for ease of V1 through standardisation) and the ALFIE performed comparably to an established measure in categorising objective memory performance – showing convergent validity. Moreover, the ALFIE showed stronger relationships with subjective appraisal of memory – indicating potentially greater concurrent and ecological validity, in comparison with the established measure. In addition, the particular administration procedure used to capture ALF with the ALFIE appeared to be more acceptable/ practical (with low attrition rates).

4.1 Reliability

4.1.1 Parallel forms reliability

Results suggested poor parallel forms reliability of ALFIE forms V1 and V2. There are several ways this result could be interpreted. It is possible that the ALFIE forms are not true parallel forms due to differences in difficulty level of the material. Gatewood, Feild and Barrick (2011) suggest that true form equivalency is very hard to create, due to difficulty realistically controlling for all item differences. Therefore, the term *alternate forms*, might be a better descriptor than parallel forms (Hunter & Schmidt, 2004) and this applies to the ALFIE forms, as both versions do show equivalency to a current objective memory measure and thus are approximate forms, even if they do not meet the criteria of parallel forms. Alternatively, the forms may be parallel, with test performance being mediated by uncontrolled-for confounding variables. For example, Small (2002) suggested that factors such as diet, physical exercise and stress levels impact on test performance. Furthermore, it is well documented that mood disturbances, such

as increased anxiety or depression, impact on memory ability (Bell & Giovagnoli, 2007; Butler & Zeman, 2008; Chepenik, Cornew & Farah, 2007) (see Extended paper 4.1 for a discussion of the relationship between these potentially confounding variables and objective memory performance). However, exclusion criteria and counterbalancing of the sample did attempt to minimise variables such as this. Although there is the possibility that exclusion criteria would not have excluded sub-clinical samples of, for example, those low in mood.

It is possible that linear equating (or equipercentile equating if a larger sample size were to be recruited) could be conducted to compare scores if a participant was re-tested on the ALFIE and thus had completed both V1 and V2. Equating would allow a clinician to directly compare participant performance on alternate forms, despite any difference that may be present in test version difficulty. However, once we adjusted for the ease of V1 through standardisation, we propose that the forms can be used as alternate forms.

4.1.2 Inter-rater reliability

Assessment of inter-rater reliability is important as it is likely that there will be a variety of professionals scoring data in a clinical or research setting, who may interpret responses differently. McHugh (2012) proposed that kappa values ≥.60 in healthcare and clinical settings would suggest adequate inter-rater reliability of the measure. Results in the current study suggested that independent raters would appraise responses similarly for both ALFIE V1 and V2 test forms, with the kappa value of .99 indicating 'almost perfect agreement' (as determined by Cohen's [1960] qualitative description of values).

Inter-rater reliability does not appear to be commonly reported in the ALF literature when extending tests to longer-delays (see e.g., Gascoigne et al. 2014; Muhlert et al., 2011). This may be due to researchers, on the whole, utilising already published measures and therefore assuming that inter-rater reliability has already been examined. However, when reviewing the literature for those who created new ALF memory measures we experienced a similar difficulty, with a paucity of inter-rater reliability analysis (see e.g., Muhlert et al., 2010). We therefore compared our value to that found of the closest 'equivalent' test from the BMIPB, the BMIPB story. The BMIPB story showed high inter-rater reliability (r = .90) (Coughlan et al., 2007). It therefore appears that the ALFIE is in-line with

this current ‘objective’ memory measure (being slightly, but not significantly, more reliable when scoring between raters).

4.2 Validity

4.2.1 Convergent validity

Convergence of percentiles obtained by participants on recall tasks from the ALFIE and recall tasks from the BMIPB was observed (both with story tasks and the word list task). This suggests that the ALFIE provides an accurate test of memory performance over time and can provide a valid measure of memory as it performs similarly to an established, validated measure.

Our results regarding convergence of story tasks with word list tasks fits with existing research (see e.g., Delis, Cullum, Butters, Carins & Prifitera, 1988; Neblina, 2012). However, recent findings suggest that outside of a non-clinical sample task types should not be taken as interchangeable as abilities in other cognitive domains can impact on memory performance differently dependent on task (Helmstaedter et al., 2009; Wicklund et al., 2006; Zahodne et al., 2011). However, despite this Helmstaedter et al. (2009) were still able to conclude that memory test performance on both task types could distinguish between epilepsy types, therefore, it is important to consider what the test is being administered for.

Convergence was not seen with recognition tasks on the ALFIE and the BMIPB. It may be that the impact of other cognitive domains on memory performance (mentioned above) has a greater impact of recognition over recall tasks, however the literature examining this has focused mainly on recall and so we cannot draw any firm conclusions about this.

4.2.2 Ecological validity

As convergence was shown between the BMIPB and the ALFIE it is possible that testing intervals of the existing measure could be adapted, as normative data could be utilised from this study. However, although extension of the published measure to a two-week delay might seem justifiable, correlations between the ALFIE and the subjective memory measure showed larger effect sizes than the published story and word lists tasks. Effect sizes in the literature between subjective and objective memory measures have been found to be moderate (.3

– .4) (Brown et al., 1991; O'Shea et al., 1996). In the present study, although several sub-sections of the ALFIE did present with a similarly moderate effect size (40%), 20% showed a large effect size (largest correlation = .8).

Furthermore, where moderate-large effect sizes were seen, they involved correlations between subjective memory scores and difference scores involving the T3 time-point (i.e. either difference in recall or recognition between T1-T3 or T2-T3). We propose that this is suggestive of the ALFIE being a more veridical measure than established objective memory measures and that it may be useful for indicating longer term forgetting or ALF-related difficulties, due to the relation with T3 difference scores. Larger effect sizes with the EMQ indicate greater ecological validity of the ALFIE measure as it more closely approximates the real-life situation experienced by the participants. Test veridicality is extremely important for a population experiencing ALF as past research has noted a discrepancy between subjective and objective memory scores in this population, with objective memory measures not finding the same level of memory difficulty as described by the individual (see e.g. Baños et al., 2004; Piazzini et al., 2001; Thompson & Corcoran, 1992), suggesting that the objective memory measures are not convergent with the individual's real-life situations.

However, it should be noted that effect sizes with Q29 of the EMQ (that specifically addressed ALF-type memory difficulties) were similar across the ALFIE and the BMIPB. It may be that this question did not fully encapsulate the difficulties that participants experienced with regards to long-term forgetting and that use of the overall measure may be more encompassing. Furthermore, as it was expected that those in our normative sample would not experience ALF it would be expected for correlations to be more prominent across their everyday memory experience. Examination with a clinical sample will be helpful, as those with 'healthy' memory functioning are suggested to perceive their abilities differently to those with poorer memories (Helmstaedter et al., 1998).

We feel that this highlights the importance of using real-life measures to mimic day-to-day experiences. Increased veridicality may also be a result of multi-modal presentation methods. This does fit with other ALF research that utilises real-life, multi-modal stimuli (see e.g., Muhlert et al., 2010), however these measures were not clinically practical, something we feel the ALFIE has addressed.

4.3 Clinical practicability

In addition to examining the ecologically validity of the test materials, we also aimed to examine the clinical practicability of the measure. This issue is particularly pertinent to the assessment of memory in Epilepsy, given that Epilepsy clinics are often specialist services covering large geographical areas. This means it would be impractical and expensive for clients to return to clinic for a two-week memory assessment follow-up, which may only last 15 minutes. Furthermore, it may be difficult to implement, asking clients to attend two memory assessment sessions exactly two weeks (±2 days) apart. The ALFIE addressed this by conducting the second assessment session remotely over the telephone. We had a low attrition rate, with only two out of 53 participants being non-contactable at two-week follow-up (3.8%). This attrition rate may be lower in a clinical population who would be more invested in the memory assessment process. Corbett's (2012) research concurs with this notion, as only one out of 44 participants were non-contactable at two-week follow-up (a 2.27% attrition rate). Attrition rates for the ALFIE appear to be comparatively low when compared to similar research into ALF-related difficulties, but that conducted face-to-face extended-delay follow-ups, for example Blake et al. (2000) found that 15.79% of their control sample did not attend the eight-week follow-up (and 8.70% of their clinical sample) and this was despite the delay coinciding with a concurrent neurology clinic. Although the low attrition rate cannot be attributed solely to qualities of the ALFIE (as the rate of people dropping out could be due to, e.g. either the BMIPB or the ALFIE), this increased retention rate does suggest that use of two-week delayed assessment via telephone might be a clinically viable solution for assessing ALF-related memory complaints (see Extended section 4.2).

4.4 Creation of a standardisation sample

A standardised sample was created to enable provision of norms for use by clinicians and researchers.

4.4.1 Age

Participants were recruited on the basis of pre-defined age categories based on the WAIS (see e.g., Wechsler, 1997b, 2008), as the general memory literature suggests that increasing age impacts negatively on memory ability (see e.g., Old & Naveh-Benjamin, 2008; Zacks, Hasher & Li, 2000). However, we found no

significant effect of age on objective memory test performance. This pattern of performance is atypical as explicit tasks (such as recall and recognition tasks used for objective memory assessment) are suggested to be more negatively impacted on by age than implicit tasks (see e.g., Davis, Trussell, & Klebe, 2001; Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004; Wilson, Leurgans, Boyle, & Bennett, 2011), as is memory for contextual information (for example, the ALFIE visual recognition task) over narrative recall and recognition (Simons, Dodson, Bell, & Schacter, 2004).

However, ours is not the first study to have found rates of forgetting comparable across ages (e.g. Hess, Auman, Colcomber, & Rahhal, 2003; Old & Naveh-Benjamin, 2008; Rahhal, Hasher, & Colcombe, 2001), although these studies highlight that comparability is mediated by variables such as the type of information being presented or the perception of test context. For example, age-related deficits can be minimised when information being presented is of a positive nature (Old & Naveh-Benjamin, 2008). Furthermore, previous studies have found that if stereotypical age-related memory deficits are highlighted by researchers (e.g., through wording of task instructions), this can be viewed as a threat and impede memory performance, particularly with increasing age, as participants are concerned that they will be judged based on an age stereotype (see e.g., Hess et al., 2003; Rahhal et al., 2001). In our study, stereotype threat may have been perceived as low, as the focus on the PIS was on creation of norms to help an epilepsy population, as opposed to drawing attention to any age-related memory difficulties. Alternately, it may be that the recruitment process attracted those adults in the older age brackets who felt confident about their memory abilities and held a positive image of their aging self. Positive age stereotypes do not negatively impact on memory performance and have even been shown to improve memory abilities (Hess et al., 2003; O'Brien & Hummert, 2006).

4.4.2 Gender

Our study indicated that females performed better on V1 of the ALFIE than males. This finding is mirrored in many previous memory studies showing that females perform better on several tasks: word recall (Bolla-Wilson & Bleecker, 1986), word recognition (Temple & Cornish, 1993), story recall (Hultsch, Masson & Small, 1993), picture recall (Galea & Kimura, 1991) and odour recognition

(Lehrner, 1993). Although there is some conflict within the literature (see e.g., Grysman & Hudson, 2013) a large body of research states that females feel more emotional intensity in connection with materials and therefore recollect more emotional details and include more narrative information during memory recall, which increases their performance on autobiographical and episodic tasks, as compared to males (Ely & Ryan, 2008; Pillemer, Wink, DiDonato & Sanborn, 2003). Although both ALFIE stories V1 and V2 have reference to emotions and internal states of the person in the narrative, V1 utilises more emotive language (for example, in V1: Roy Wyer is “absolutely surprised and gob-smacked” and in V2: Sarah Outen experienced “highs and lows”). Furthermore, the V1 story appeared to have more resonance with participants, as determined by the primary author’s observations (for example laughter, exhales of breathe, ‘huffs’ at certain story points, comments during and after video clip presentation). This may have resulted in females feeling a heightened emotional resonance with this video-clip and thus performed better on this version.

Grysman and Hudson (2013) suggest that gender differences may or may not be seen dependent on study methodology and context and it was felt important to examine these in reference to the current study. They suggest that small sample sizes, disproportionate gender ratios and choice of materials, such as the use of rating scales, can affect study outcomes. For example, studies utilising rating scales to examine emotional intensity felt after presentation of a narrative autobiographical memory have often found no indication of gender differences (see e.g., Escobedo & Adolphs, 2010, Neumann & Phillipot, 2007). However, when examined, there are methodological issues such as skewed gender ratios (e.g. 13 males and 45 females in Neumann & Phillipot’s [2007] study). Furthermore, it may be that this more explicit measure of emotional resonance with information does not provide adequate questions to tap into gender differences. In the current study the most pertinent limitation in relation to Grysman and Hudson’s (2013) critique of the literature in this area would be the small sample size afforded to this study.

Finally, as a contextual factor, it has been argued that males do not emotionally elaborate during recall unless it is felt to be relevant to the situation, whereas females do so more freely (Fivush, Bohanek, & Zaman, 2011) and will provide more specific details than males (Pillemer et al., 2003). This may mean that male

participants would have benefited from a prompt to elaborate on the emotional content of the story. This may have been more pertinent to ALFIE V1, where the story had more highly emotional content and connection with participants (as previously discussed).

4.4.3 Stability of memory

A significant association between performance on the ALFIE at all time points was found for V2 participants. However, relationships between memory performance on the ALFIE at different time points varied for V1 participants according to gender (See section 4.4.2 for possible effect of gender). We expected some degree of inter-dependence (as all rely on initial learning) but we felt it might be clinically useful to know the 'normal' level of correlation, as for example, in those experiencing ALF, these associations might be unexpectedly (relatively) low between T2 and T3.

4.4.4 Clinical population

Use of the newly-created ALFIE standardised norms was used to examine a clinical TLE sample from Corbett's (2012) research. There was variability in the ability of the ALFIE to categorise these participants as experiencing abnormal rates of forgetting. However, the problem with Corbett's (2012) study is that it is unclear whether any of the participants were actually complaining of ALF, with most seeming to be fairly happy with their memory (as determined by the MFQ), which makes it hard to say what we expect the ALFIE to be doing in this group. It may have been more useful to have a group who were subjectively reporting impairment/ seeking treatment for ALF, then we could say more about the ability of the ALFIE to detect problems (or at least whether the two-week recall is more useful than the 40-minute recall).

Corbett's (2012) data could be split into those that did subjectively report memory problems (n = 6) (based on a cut-off of <3.5 on frequency of forgetting on the MFQ [where lower scores indicate greater problems]). However, the difficulty here was that for one of these TLE participants, the observer disagreed by >4 points (the observer saw no memory problems), and furthermore, as very few of Corbett's (2012) participants subjectively reported memory problems, and then these participants had to be split into their relevant test versions and gender, a larger clinical sample would be needed to draw any firm conclusions.

4.5 Limitations and future directions

We only recruited 53 participants to the study, with only 50 participants being included in all analyses. Due to splitting the data into V1 and V2 and then subsequently some sub-tests by gender, sample sizes for analysis were small. A larger sample size would be desirable for future studies as it would afford greater power and precision of effect detection, and thus a reduction in potential for Type I errors in (particularly mixed ANOVA) results, meaning significant effects of age group and years of education may remain after post-hoc analysis.

Although there was no significant effect of age on objective memory test performance, mixed ANOVAs examining this effect on ALFIE Story V2 scores were statistically significant before Bonferroni corrections were applied. Cardinal and Aitken (2006) state that this is common when dealing with small sample sizes. It is unclear whether with a larger sample size this effect may have remained significant. A larger sample would clarify whether this effect reflects (1) a genuine Type I error or (2) a lack of power. If the observed effect size is accurate (i.e., is of a similar magnitude in a larger sample), then it would likely reach significance in a larger sample.

As Grysman and Hudson (2013) state, many of the studies reporting an effect of gender on memory performance were not reviewing this as a primary aim of their research (similar to the current study). Examining the gender difference observed across some of the objective memory measures as a primary aim, rather than a peripheral factor, may be beneficial to fully allow examination of this effect as it would specifically power the study (future studies) to detect gender effects (if present). It may also be useful to provide a measure of the emotional impact of ALFIE story V1 and V2 on participants to further examine whether this is a mediating factor for the gender difference noted. Interviewing participants about the emotional salience of ALFIE V1 and V2 may also help to pinpoint whether test versions were experienced differently.

In addition, the ALFIE utilises a 0-1 marking system as opposed to the established objective measure, which has a 0-1-2 marking system. A 0-1 marking system means that scoring becomes more stringent and participants need to remember the narrative as close to the words that were said as possible in order to achieve marks. In relation to research suggesting that females emotionally

elaborate more freely (Fivush et al., 2011) and provide more specific details than males in recall (Pillemer et al., 2003) it was felt that the marking criteria may have therefore disadvantaged male participants. The development of a 0-1-2 scoring system may therefore be a useful future direction. Although, ultimately we produced gender-specific norms to mitigate this. Any individual factors that influence performance could potentially be measured and normed for in the same manner.

Finally, although the ALFIE was developed with a TLE population in mind, ALF may be experienced more broadly (and conversely, will not be experienced by all individuals with TLE). The general set of norms here could potentially be used/ developed further with any adults experiencing ALF. However, examination of cultural differences within English-speaking demographics would be useful to determine replicability 'within-language'. (See Extended section 4.3 for further information on the limitations/ directions for future research.)

4.6 Conclusion

The ALFIE test shows good inter-rater reliability, convergent validity and ecological validity. Although the parallel forms reliability of V1 and V2 has to be questioned due to results indicating that forms may not be parallel, research indicates that true parallel forms are unlikely, and future research examining confounding variables and allowing for larger sample sizes may reduce this concern somewhat. Furthermore, through standardisation ease of test version can be accounted for and the versions may be used as alternate forms. Test veridicality indicates that the ALFIE may be a better predictor of real-life memory difficulties than current objective memory measures; something that is particularly salient in a TLE population where individuals going to memory clinics find a 'mis-match' between their subjective accounts and scores on formal objective measures. Furthermore, the low attrition rate of participants suggests that a remote telephone method for delayed two-week assessment is a viable method to adopt within a clinical context.

Conflict of interest

There is no conflict of interest to declare by the authors.

Role of funding source

The University of Lincoln provided financial support for this research study, but had no involvement in the study design, data collection, analysis or interpretation, or in the writing of this report.

Acknowledgements

Thanks go to all of the participants who gave their time to take part in this research.

References

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences: Neurobiology, 91,* 7041-7045.

Andelman, F., Zuckerman-Feldhay, E., Hoffien, D., Fried, I., & Neufeld, M. Y. (2004). Lateralisation of deficit in self-awareness of memory in patients with intractable epilepsy. *Epilepsia, 45(7),* 826-833.

Annett, J., McLaughlin Cook, N., & Leslie, J. (1995). Interference with olfactory memory by visual and verbal tasks. *Perceptual and Motor Skills, 80,* 1307-1317.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. II, pp. 89-195). New York, USA: Academic Press.

Baddeley, A., Eysenck, M.W., & Anderson, M.C. (2009). *Memory.* London, UK: Psychology Press.

Baddeley, A., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short- term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176-189.

Baños, J. H., LaGory, J., Sawrie, S., Faught, E., Knowlton, R., Prasad, A., Kuzniecky, R., & Martin, R. C. (2004). Self-report of cognitive abilities in temporal lobe epilepsy: cognitive, psychosocial, and emotional factors. *Epilepsy & Behavior, 5(4),* 575-579. DOI: 10.1016/j.yebeh.2004.04.010

Bell, B. D. (2006). WMS-III logical memory performance after a two-week delay in temporal lobe epilepsy and control groups. *Journal of Clinical & Experimental Neuropsychology, 28,* 1435-1443.

Bell, B. D., & Giovagnoli, A. R. (2007). Recent innovative studies of memory in temporal lobe epilepsy. *Neuropsychology Review, 17,* 455-476.

Bigelow, J., & Poremba, A. (2014). Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. *PLoS ONE, 9(2),* e89914. DOI: 10.1371/journal.pone.0089914

Blake, R. V., Wroe, S. J., Breen, E. K., & McCarthy, R. A. (2000). Accelerated forgetting in patients with Epilepsy. *Brain, 123,* 472-483.

Bolla-Wilson, K., & Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age, and education on the Rey Auditory Verbal Learning Test. *Developmental Neuropsychology, 2,* 203–211.

Brown, F. H., Dodrill, C. B., Clark, T., & Zych, K. (1991). An investigation of the relationship between self-report of memory functioning and memory test performance. *Journal of Clinical Psychology, 47,* 772-777.

Browne, T. R., & Holmes, G. L. (2008). *Handbook of Epilepsy.* (4th ed.). Philadelphia, PA: Lippincott Williams & Wilkins (LWW).

Butler, C. R., & Zeman, A. Z. (2008). Recent insights into the impairment of memory in epilepsy: Transient epileptic amnesia, accelerated long-term forgetting and remote memory impairment. *Brain, 131,* 2243-2263. DOI: http://dx.doi.org/10.1093/brain/awn127

Cardinal, R. N., & Aitken, M. R. F. (2006). *ANOVA for the behavioural sciences researcher.* New Jersey, USA: Lawrence Erlbaum Associates.

Chalmers, A. (1999). *What is this thing called science?* (3rd ed.). Maidenhead, UK: Open University Press.

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion, 7(4),* 802-811. DOI: 10.1037/1528-3542.7.4.802

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. Education and Psychological Measurement, 20, 37–46.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112(1),* 155–159. DOI: 10.1037/0033-2909.112.1.155

Corbett, F. (2012). The challenge of assessing accelerated long-term forgetting in Temporal Lobe Epilepsy: A study addressing issues of ecological-validity and clinical practicality. (Unpublished doctoral dissertation). Nottingham, UK: University of Nottingham.

Cornish, I. M. (2000). Factor structure of the Everyday Memory Questionnaire. *British Journal of Psychology, 91(3),* 427-438. DOI: 10.1348/000712600161916

Coughlan, A. K., Oddy, M., & Crawford, J. (2007). *BIRT Memory and Information Processing Battery*. Horsham, UK: Brain Injury Rehabilitation Trust (BIRT).

Davis, H. P., Trussell, L. H., & Klebe, K. J. (2001). A ten-year longitudinal examination of repetition priming, incidental recall, free recall, and recognition in young and elderly. *Brain and Cognition, 46(1-2),* 99-104.

Delaney, R. C., Rosen, A. J., Mattson, R. H., & Novelly, R. A. (1980). Memory function in focal epilepsy: A comparison of non-surgical, unilateral temporal lobe and frontal lobe samples. *Cortex, 16(1),* 103-117.

Delis, D. C., Cullum, M., Butters, N., Cairns, P., & Prifitera, A. (1988). Wechsler Memory Scale – Revised and California Verbal Learning Test: convergence and divergence. *The Clinical Neuropsychologist, 2,* 188-196.

Escobedo, J. R., & Adolphs, R. (2010). Becoming a better person: Temporal remoteness biases autobiographical memories for moral events. *Emotion, 10,* 511–518.

Efklides, A., Yiultsi, E., Kangellidou, T., Kounti, F., Dina, F., & Tsolaki, M. (2002). Wechsler Memory Scale, Rivermead Behavioural Memory Test, and Everyday Memory Questionnaire in healthy adults and Alzheimer's patients. *European Journal of Psychological Assessment, 18,* 63–77.

Ely, R., & Ryan, E. (2008). Remembering talk: Individual and gender differences in reported speech. *Memory, 16,* 395–409.

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, H., Elger, C. E., Engel Jr, J., Forsgren, L., French, J. A., Glynn, M., Hesdorffer, D. C., Lee, B. I., Mathern, G. W., Moshé, S. L., Perucca, E., Scheffer, I. E., Tomson, T., Watanabe, M., & Wiebe, S. (2014). A practical clinical definition of epilepsy. *Epilepsia, 55(4),* 475-482. DOI: 10.1111/epi.12550

Fitzgerald, Z., Thayer, Z., Mohamed, A., & Miller, L. (2013). Examining factors related to accelerated long-term forgetting in epilepsy using ambulatory EEG monitoring. *Epilepsia, 54(5),* 819-827.

Fivush, R., Bohanek, J. G., & Zaman, W. (2011). Personal and intergenerational narratives in relation to adolescents' well-being. In T. Habermas (Ed.). *The development of autobiographical reasoning in adolescence and beyond* (pp. 45-58). San Francisco: Jossey-Bass.

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19(4),* 617-625. DOI: http://dx.doi.org/10.1037/0882-7974.19.4.617

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14,* 53–65.

Gallace, A., & Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychological Bulletin, 135(3),* 380-406. DOI: http://dx.doi.org/10.1037/a0015325

Gascoigne, M. B., Smith, M. L., Barton, B., Webster, R., Gill, D., & Lah, S. (2014). Accelerated long-term forgetting in children with temporal lobe epilepsy. *Neuropsychologia, 59,* 93-102. DOI: 10.1016/j.neuropsychologia.2014.04.012

Gatewood, R., Feild, H., & Barrick, M. (2011). *Human resource selection* (6th ed.). Mason, OH: South-Western.

Giddings, L., & Grant, B. (2007). A trojan horse for positivism?: A critique of mixed methods research. *Advances in Nursing Science, 30(1),* 52-60.

Gilewski, M. J., & Zelinski, E. M. (1988). The Memory Functioning Questionnaire. *Psychopharmacology Bulletin, 24(4),* 665-670.

Gilewski, M. J., Zelinski, E. M, and Schaie, K. W. (1990). The Memory Functioning Questionnaire. *Psychology and Aging, 5(4),* 482-490.

Giovagnoli, A. R., & Avanzini, G. (1999). Learning and memory impairment in patients with temporal lone epilepsy: Relation to the presence, type, and location of brain lesion. *Epilepsia, 40,* 904-911.

Granacher Jr., R. P. (2008). *Traumatic brain injury: Methods for clinical and forensic neuropsychiatric assessment.* (2nd ed.). Florida, USA: CRC Press.

Grysman, A., & Hudson, J. A. (2013). Gender differences in autobiographical memory: Developmental and methodological considerations. *Developmental Review, 33,* 239-272. DOI: http://dx.doi.org/10.1016/j.dr.2013.07.004

Hall, K. E., Isaac, C. L., & Harris, P. (2009). Memory complaints in Epilepsy: An accurate reflection of memory impairment or an indicator of poor adjustment? A review of the literature. *Clinical Psychology Review, 29,* 354-367.

Helmstaedter, C., Hauff, M., & Elger, C. E. (1998). Ecological validity of list-learning tests and self-reported memory in healthy individuals and those with temporal lobe epilepsy. *Journal of Clinical & Experimental Neuropsychology, 20(3),* 365-75. DOI: 10.1076/jcen.20.3.365.824

Helmstaedter, C., Wietzke, J., & Lutz, M. T. (2009). Unique and shared validity of the "Wechsler logical memory test", the "California verbal learning test", and

the "verbal learning and memory test" in patients with epilepsy. *Epilepsy Research, 87*(2-3), 203-212.

Hendriks, M. P., Aldenkamp, A. P., van der Vlugt, H., Alpherts, W. C., & Vermeulen, J. (2002). Memory complaints in medically refractory epilepsy: Relationship to epilepsy-related factors. *Epilepsy & Behavior, 3(2),* 165-172. DOI: http://dx.doi.org/10.1006/ebeh.2002.0320

Hermann, B. P., Seidenberg, M., Bell, B., Woodard, A., Rutecki, P., & Sheth, R. (2000). Comorbid psychiatric symptoms in temporal lobe epilepsy: Association with chronicity of epilepsy and impact on quality of life. *Epilepsy & Behavior, 1(3),* 184-90.

Herrman, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin, 92(2),* 434-452.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *Journal of Gerontology, 58B(1),* 3-11.

Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in Multiple Sclerosis. *Archives of Clinical Neuropsychology, 15(3),* 185-204. DOI: 10.1016/S0887-6177(99)00004-9

Hoefeijzers, S., Dewar, M., Della Sala, S., Zeman, A., & Butler, C. (2013). Accelerated long-term forgetting in transient epileptic amnesia: an acquisition or consolidation deficit? *Neuropsychologia, 51(8),* 1549-1555. DOI: 10.1016/j.neuropsychologia.2013.04.017

Hultsch, D. F., Masson, M. E., & Small, B. J. (1991). Adult age differences in direct and indirect tests of memory. *Journals of Gerontology, 46,* 22-30.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. (2nd ed.). Newbury Park, California: Sage Publications, Inc.

Jansari, A. S., Davis, K., McGibbon, T., Firminger, S., & Kapur, N. (2010). When 'long-term memory' no longer means 'forever': Analysis of accelerated long-term forgetting in a patient with temporal lobe epilepsy. *Neuropsychologia, 48(6),* 1707-1715.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Lehrner, J. P. (1993). Gender differences in long-term odor recognition memory: Verbal versus sensory influences and the consistency of label use. *Chemical Senses, 18,* 17–26.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282. DOI: http://dx.doi.org/10.11613/BM.2012.031

Millon, T., Millon, C., Davis, R., & Grossman, S. (2009). *MCMI-III Manual* (4th ed.). Minneapolis, MN: Pearson Education Inc.

Muhlert, N., Milton, F., Butler, C. R., Kapur, N., & Zeman, A. Z. (2010). Accelerated forgetting of real-life events in transient epileptic amnesia. *Neuropsychologia, 48(11),* 3235-3244.

Muhlert, N., Grünewald, R. A., Hunkin, N. M. Reuber, M., Howell, S., Reynders, H., & Isaac, C.L. (2011). Accelerated long-term forgetting in temporal lobe bt not idiopathic generalised epilepsy. Neuropsychologia, 49(9), 2417-2426.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7(2),* 217-227.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000), Multiple trace theory of human memory: Computational, neuroimaging, and

neuropsychological results. *Hippocampus, 10(4),* 352–368. DOI: 10.1002/1098-1063

National Institute for Health and Clinical Excellence (NICE). (2012). *The epilepsies: The diagnosis and management of the epilepsies in adult and children in primary and secondary care.* London, UK: Author.

Neblina, C. (2012). Construct and criterion validity of the Rey Auditory Verbal Learning Test-Spanish Version in adults with traumatic brain injury. (Unpublished masters dissertation). Las Vegas, Nevada: University of Nevada.

Neumann, A., & Phillipot, P. (2007). Specifying what makes a personal memory unique enhances emotion regulation. *Emotion, 7,* 566–578.

O'Brien, L. T., & Hummert, M. L. (2006). Self-stereotyping, stereotyping threat, and memory performance in middle-aged adults. *Social Cognition*, *24,* 338-358.

Old, S., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23,* 104-118.

The Organisation for Economic Co-operation and Development (OECD). (2014). Education at a Glance 2014: OECD indicators. n.p.: Authors. DOI: http://dx.doi.org/10.1787/eag-2014-en

O'Shea, M. F., Saling, M. M., Bladin, P. F., & Berkovic, S. F. (1996). Does naming contribute to memory self-report in temporal lobe epilepsy? *Journal of Clinical & Experimental Neuropsychology, 18(1),* 98-109. DOI: 10.1080/01688639608408266

Piazzini, A., Canevini, M. P., Maggiori, G., & Canger, R. (2001). The perception of memory failures in patients with epilepsy. *European Journal of Neurology, 8(6),* 613-20. DOI: 10.1046/j.1468-1331.2001.00287.x

Pillemer, D. B., Wink, P., DiDonato, T. E., & Sanborn, R. L. (2003). Gender differences in autobiographical memory styles of older adults. *Memory, 11,* 525–532.

Ponds, R. W., & Hendriks, M. (2006). Cognitive rehabilitation of memory problems in patients with epilepsy. *Seizure, 15(4),* 267-273. DOI: http://dx.doi.org/10.1016/j.seizure.2006.02.011

Rahhal, T. A., Hasher, L., & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging, 16*, 697–706.

Royle, J., & Lincoln, N. B. (2008). The Everyday Memory Questionnaire-Revised: development of a 13-item scale. *Disability & Rehabilitation, 30(2),* 114-121. DOI: 10.1080/09638280701223876

Sbordone, R. J., & Long, C. J. (Eds.). (1996). *Ecological validity of neuropsychological testing*. Boca Raton, FL: St Lucie Press.

Simons, J. S., Dodson, C. S., Bell, D., & Schacter, D. L. (2004). Specific- and partial-source memory: effects of aging. *Psychology and Aging, 19(4),* 689–694. DOI: http://dx.doi.org/10.1037/0882-7974.19.4.689

Small, G. W. (2002). What we need to know about age related memory loss. *British Medical Journal, 324(7352)*, 1502-1505.

Squire, L.R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology, 5,* 169-l77.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience, 27(1),* 279–306. DOI: 10.1146/annurev.neuro.27.070203.144130

Sunderland, A., Harris, J. E., & Baddeley, A. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22,* 341-357.

Sunderland, A., Harris, J. E., & Gleave, J. (1984). Memory failures in everyday life following severe head injury. *Journal of Clinical Neuropsychology, 6*(2), 127-142. DOI: 10.1080/01688638408401204

Temple, C. M., & Cornish, K. M. (1993). Recognition memory for words and faces in schoolchildren: A female advantage for words. *British Journal of Developmental Psychology, 11,* 421–426.

Thompson, P., & Corcoran, R. (1992). Everyday memory failures in people with epilepsy. *Epilepsia, 33(6),* 18-20.

Tombaugh, T. N. (1996). *Test of Memory Malingering manual*. New York, USA: MultiHealth Systems.

Tombaugh, T. N. (1997). The Test of Memory Malingering (ToMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*, 260–68.

Viskontas, I. V., McAdnres, M. P., & Moscovitch, M. (2000). Remote episodic memory deficits in patients with unilateral temporal lobe epilepsy and excisions. *The Journal of Neuroscience, 20(15),* 5853-5857.

Wechsler, D. (1997a). *Wechsler Memory Scale-Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Adult Intelligence Scale-Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading.* San Antonio, TX: The Psychological Corporation.

Wicklund, A. H., Johnson, N., Rademaker, A., & Weitner, B. B., & Weintraub, S. (2006). Word list versus story memory in Alzheimer disease and

frontotemporal dementia. *Alzheimer's Disease and Associated Disorders, 20(2),* 86-92.

Wilson, R. S., Leurgans, S. E., Boyle, P. A., & Bennett, D. A. (2011). Cognitive decline in prodromal Alzheimer disease and mild cognitive impairment. *Arch Neurol, 68(3),* 351-6.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In T. A. Salthouse & F. I. M. Craik (Eds.), *Handbook of Aging and Cognition* (2nd ed.). (pp. 293-357). Mahwah, NJ: Lawrence Erlbaum.

Zahodne, L. B., Bowers, D., Price, C. C., Bauer, R. M., Nisenzon, A., Foote, K. D., & Okun, M. S. (2011). The case for testing memory with both stories and word lists prior to DBS surgery for Parkinson's disease. *Clin Neuropsychol, 25(3),* 348-358. DOI: 10.1080/13854046.2011.562869

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67(6),* 361-370.

Appendices

Appendix C. ALFIE Story recall score sheets

ALFIE V1 Story recall score sheet

Roy / Wyer / was fined £75 / for littering / after brushing his dog's hair / in a park / in Bilborough, near the Harvey Haddon Stadium. / It followed a disagreement with a community protection officer / who'd accused him of impersonating an officer. / Quentin Rayner reports. / The trouble started when Roy Wyer was spotted brushing / his dog Spencer in a park near the Harvey Haddon stadium by a community protection officer. / Initially the CPO objected to the sergeant's chevron's / on Roy's high visibility jacket, / which he'd earned from his days as a security guard. / The CPO questioned whether he was impersonating an officer. / When Roy refused to give his details the CPO issued him with a £75 fine for littering the park with large clumps of dog fur. / I was absolutely gob-smacked, absolutely surprised and gob-smacked. I couldn't believe what I was hearing. And this is all they done me for [reporter: that's all]… that's all the hair that came off the dog…[reporter: and that's what you were fined £75 for?] / Yes that's right, that's correct…and if I don't pay it at all it'll cost me £2000 pounds / and-er- I'll be arrested / the city council has started an enquiry and / the CPO involved may get retrained. / The fine was immediately overturned / and Roy has been issued with an apology. / Well I've not received it yet but-erm-when it's writing I'll probably believe it, so I'll just wait and see. / Roy thinks it's a case of people going power mad / and has every intention of continuing to brush Spencer's coat in the park. / Quentin Rayner, East Midlands today, Bilborough.

Total score = _____/28

ALFIE V2 Story recall score sheet

News from Sarah / Outen. / The 26 year old / has completed the tricky first leg of her record breaking human powered / loop of the globe. / Two years ago Sarah was awarded an MBE / after becoming the first woman to row solo across the Indian Ocean. / Today she arrived in Tokyo after an epic expedition across land and sea, / Sarah Teal reports. / Since setting off from London / on April the 1st / Sarah Outen has covered a lot of land and sea / on her bike and in a kayak. / Seven months on she's reached Tokyo. It's the end of the first leg of her London to London, record breaking attempt to loop the globe. / It's been full of adventures and challenges and highs and lows and / I suppose there were times when I wasn't quite sure how I was going to make it this far…um… / whether that's boshing through the heat of the Gobi desert / or in thick mud in Russia, / bits going wrong with bits of equipment and so on…it's been brilliant. / Sarah from Rutland in Oakham / has enjoyed the highs of the beautiful landscape / and the local wildlife – / I'm about fifty metres away from a brown bear [inaudible 'breathe in'] / – and endured the lows of the dangerous roads / and treacherous seas. / Since leaving London Sarah has travelled 11,000 miles, / through 12 countries / and kayaked 300 nautical miles to reach Japan.

Total score = _____/28

Appendix D. ALFIE Story recall scoring guidelines

ALFIE V1 Story recall scoring guidelines

	Information unit	Words with the same or a close meaning are acceptable…e.g.
1	Roy	-
2	Wyer	-
3	was fined £75	Charged/ made to pay
4	for littering	Dumping/ throwing away/ disposing of
5	after brushing his dog's hair	Fur/ coat/ clippings
6	in a park	Playing field
7	in Bilborough OR near the Harvey Haddon Stadium.	-
8	It followed a disagreement with a community protection officer	-
9	who'd accused him of impersonating an officer.	Mimicking/ pretending to be Policeman
10	Quentin Rayner reports.	-
11	The trouble started when Roy Wyer was spotted brushing	Grooming/ combing
12	his dog Spencer in a park near the Harvey Haddon stadium by a community protection officer.	-
13	Initially the CPO objected to the sergeant's chevron's	Lapels
14	on Roy's high visibility jacket,	Yellow/fluorescent Coat
15	which he'd earned from his days as a security guard.	Worked in security
16	The CPO questioned whether he was impersonating an officer.	-
17	When Roy refused to give his details the CPO issued him with a £75 fine for littering the park with large clumps of dog fur.	Declined/ would not/ decided not to Information/ name and address
18	I was absolutely gob-smacked, OR absolutely surprised and gob-smacked. OR I couldn't believe what I was hearing. And this is all they done me for…[reporter: that's all]… that's all the hair that came off the dog…[reporter: and that's what you were fined £75 for?]	-

19	Yes that's right, that's correct…and if I don't pay it at all it'll cost me £2000 pounds	Charge/ fine/ demand Grand/ k/ quid
20	and-er- I'll be arrested	-
21	the city council has started an enquiry and	Investigation
22	the CPO involved may get retrained.	More training
23	The fine was immediately overturned	Cancelled/ cancelled/ withdrawn
24	and Roy has been issued with an apology	They said sorry
25	Well I've not received it yet but-erm- OR when it's writing I'll probably believe it, so I'll just wait and see	-
26	Roy thinks it's a case of people going power mad	-
27	and has every intention of continuing to brush Spencer's coat in the park.	-
28	Quentin Rayner, East Midlands today, Bilborough.	-
Total score available = 28		

From Corbett (2012).

ALFIE V2 Story recall scoring guidelines

	Information unit	Words with the same or a close meaning are acceptable…e.g.
1	News from Sarah	-
2	Outen	-
3	The 26 year old	-
4	has completed the tricky first leg of her record breaking human powered	Finished/ manages/ been able to complete/ achieved…first part/ section/ bit
5	loop of the globe	Round the world trip/ expedition/ journey/ travels
6	Two years ago Sarah was awarded an MBE	
7	after becoming the first woman to row solo across the Indian Ocean	
8	Today she arrived in Tokyo after an epic expedition across land and sea,	
9	Sarah Teal reports	
10	Since setting off from London	
11	on April the 1st	
12	Sarah Outen has covered a lot of land and sea	
13	on her bike and in a kayak.	
14	Seven months on she's reached Tokyo. It's the end of the first leg of her London to London, record breaking attempt to loop the globe	
15	It's been full of adventures and challenges OR and highs and lows and	
16	I suppose there were times when I wasn't quite sure how I was going to make it this far…um…	
17	whether that's boshing through the heat of the Gobi desert	
18	or in thick mud in Russia,	
19	bits going wrong with bits of equipment and so on…it's been brilliant	
20	Sarah from Rutland in Oakham	
21	has enjoyed the highs of the beautiful landscape	
22	and the local wildlife	
23	I'm about fifty metres away from a brown bear [inaudible 'breathe in?]	
24	And endured the lows of the dangerous roads	
25	and treacherous seas	
26	Since leaving London Sarah has travelled 11,000 miles,	
27	through 12 countries	

28	and kayaked 300 nautical miles to reach Japan.
Total score available = 28	

From Corbett (2012).

Appendix E. ALFIE Recognition score sheets

ALFIE V1 Story recognition score sheet

ALFIE V1 Narrative recognition		
1	Was the man's name Roy Wyer?	**YES**/no
2	Was he fined £100?	yes/**NO**
3	Was the man's dog named Samson?	yes/**NO**
4	Was the man fined for littering clumps of dog hair?	**YES**/no
5	Was the man in the Harvey Haddon stadium?	yes/**NO**
6	Was the man approached by a community protection officer?	**YES**/no
7	Was the man accused of impersonating a police officer?	**YES**/no
8	Was the man wearing a high visibility jacket?	**YES**/no
9	Was the man gob-smacked by the accusation?	**YES**/no
10	Has the man received a written apology?	yes/**NO**
11	Will the man have to pay the fine?	yes/**NO**
12	Was the roaming reporter's name Simon Blake?	yes/**NO**
13	Does the man think that people are going power mad?	**YES**/no
14	Does the man intend to stop brushing his dog in the park?	yes/**NO**
	TOTAL CORRECT /14	

Visual feature recognition		
1	Were there two presenters in the studio?	**YES**/no
2	Was the male presenter wearing a stripy tie?	**YES**/no
3	Did the female presenter have brown hair?	yes/**NO**
4	Was the female presenter wearing a black dress?	**YES**/no
5	Did the studio presenters have mugs of tea?	yes/**NO**
6	Did the man in the story have white hair?	**YES**/no
7	Did the man in the story have a beard?	**YES**/no
8	Was the man's dog a Labrador?	yes/**NO**
9	Was the man in the story on a park bench?	yes/**NO**
10	Was the fine written on pink paper?	yes/**NO**
11	Was the man in the story wearing a yellow jacket?	**YES**/no
12	Was there a game of football happening in the park?	yes/**NO**
13	Did the roaming reporter have grey hair?	**YES**/no
14	Was the roaming reporter wearing a rain coat?	yes/**NO**
	TOTAL CORRECT /14	

ALFIE V2 Story recognition score sheet

ALFIE V2 Narrative recognition		
1	Was the woman named Sally Newton?	yes/**NO**
2	Was the woman 26 years old?	**YES**/no
3	Was the woman travelling between the north and south pole?	yes/**NO**
4	Has the woman already swum across the Indian ocean?	yes/**NO**
5	Was Tokyo the first stop on the woman's trip?	**YES**/no
6	Was the reporter's name Sarah Teal?	**YES**/no
7	Did the woman set off from London on February 14th?	yes/**NO**
8	Did it take the woman seven months to reach her first destination?	**YES**/no
9	Was the woman always certain that she would reach her destination?	yes/**NO**
10	Has the woman been through the Gobi Desert?	**YES**/no
11	Did the woman come from Grantham in Lincolnshire?	yes/**NO**
12	Did the woman see a brown bear?	**YES**/no
13	Has the woman already covered 20,000 miles?	yes/**NO**
14	Has the woman travelled through 12 countries?	**YES**/no
	TOTAL CORRECT /14	

Visual feature recognition		
1	Were there two presenters in the studio?	**YES**/no
2	Were the presenters sitting on a blue sofa?	yes/**NO**
3	Was the male presenter wearing a stripy tie?	**YES**/no
4	Did the female presenter have brown hair?	yes/**NO**
5	Was the female presenter wearing a black and white dress?	**YES**/no
6	Did the studio presenters have mugs of tea?	yes/**NO**
7	Did the woman in the story have short hair?	**YES**/no
8	Did the woman in the story have black hair?	yes/**NO**
9	Did the woman in the story wear a safety helmet on her bike?	**YES**/no
10	Did the woman have bags attached to her bike?	**YES**/no
11	Was the woman overtaken by bus when she was on her bike?	yes/**NO**
12	Did the woman wear a green life jacket in her kayak?	yes/**NO**
13	Did the woman row past a seal?	yes/**NO**
14	Was the reporter in front of the map in the studio wearing a red dress?	**YES**/no
	TOTAL CORRECT /14	

Appendix F. ALFIE percentile tables

Table 12. Percentile norms for ALFIE Story Version 1

PR	Raw scores*					
	T1		T2		T3	
	m	f	m	f	m	f
2	8	7	5	9	3	7
10		9	7	10		
25	13	12	11	13	8	8
50	15	15	14	16	10	9
75	17	18	16	17	13	11
90	19	20	18	20	15	13

*ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; m = male; f = female; PR = Percentile Range; *Rounded to the nearest whole number.*

Table 13. Percentile norms for ALFIE Narrative & Visual Recognition Version 1

PR	Raw scores*							
	Narrative T2		Visual T2		Narrative T3		Visual T3	
	m	f	m	f	m	f	m	f
2	11	11	9	7	11	9	8	5
10						10		6
25		12	10	9	12	12	9	8
50	13	13	11	10			10	9
75					14	13	11	10
90	14	14	12	11		14	12	11

*ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); T2 = 40-minute delay; T3 = 2-week delay; m = male; f = female; PR = Percentile Range; *Rounded to the nearest whole number.*

Table 14. Percentile norms for ALFIE Story Version 2

PR	Raw scores*		
	T1	T2	T3
2	4	4	0
10	5	5	1
25	6	6	2
50	7	7	4
75	10	9	6
90	11	11	8

*ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; PR = Percentile Range; *Rounded to the nearest whole number.*

Table 15. Percentile norms for ALFIE Narrative & Visual Recognition Version 2

PR	Raw scores*			
	Narrative T2	Visual T2	Narrative T3	Visual T3
2	6	7	8	7
10	9	8		8
25	10	9	9	9
50		10	10	
75	11	11	11	10
90	12	12	13	12

*ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); T2 = 40-minute delay; T3 = 2-week delay; PR = Percentile Range; *Rounded to the nearest whole number.*

Appendix G. BMIPB extended percentile tables

Table 16. Percentile norms for BMIPB tasks at T2

PR	Raw scores*							
	Recall		Recognition					
	List		List			Word		
	V1	V2	V1 m	V1 f	V2	V1 m	V1 f	V2
2	6	0	24	23	14	22	24	21
5	7	2			15			22
10	9	6	25	25	21		25	24
25	11	9	26	29	26	26	26	26
50	13	12	28		28	28	28	27
75	14	13	30	30	30	29		28
90	15	14					30	29

*BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); T2 = 40-minute delay; PR = Percentile Range; V1 = Test Version 1; V2 = Test Version 2; m = male; f = female; *Rounded to the nearest whole number.*

Table 17. Percentile norms for BMIPB tasks at T3

PR	Raw scores*								
	Recall			Recognition					
	Story	List		List			Word		
	V1 & V2	V1	V2	V1 m	V1 f	V2	V1 m	V1 f	V2
2	0	2		16	7	14			17
5	6		0			16	23	22	19
10	8	3		17	13	19			23
25	13	4	3	23	23	20	25	23	25
50	19	7	4		25	22	27	27	26
75	26	9	8	27	27	27	29	28	28
90	34	13	11	30	29	28	30	30	

*BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); T3 = 2-week delay; PR = Percentile Range; V1 = Test Version 1; V2 = Test Version 2; m = male; f = female; *Rounded to the nearest whole number.*

Extended paper

1. Extended Introduction

1.1 Memory complaints in Epilepsy

Community-based studies show around 60-70% of those with Epilepsy to subjectively complain of cognitive difficulties (Carpay, Aldenkamp, & van Donselaar, 2005; Uijl et al., 2006), with studies looking specifically at memory finding around half of the Epilepsy population highlight memory as a problem (Corcoran, & Thompson, 1992; Fisher et al., 2000). It is suggested that this percentage may be even higher in a population attending an Epilepsy clinic for care. These memory difficulties could be as a result of the condition or of condition-related factors, for example, anti-epileptic drug use (Motamedi, & Meador, 2004), structural lesions (Muhlert et al., 2011), seizure activity (Jokeit, Daamen, Zang, Jansky, & Ebner, 2001; Mameniskiene et al., 2006), or mood distrubance (Mameniskiene et al., 2006) (although it should be noted that there is some disparity in the research regarding these within-participant variables).

In one study by McAuley et al. (2010), patients were asked to rate five Epilepsy-related complaints they found the most concerning for them, out of a range of 20 items (e.g., having a seizure unexpectedly, sexual health, mood; see Epilepsy Foundation of America Concerns Index [Gilliam et al., 1999]). The third most frequently selected response was memory, which was the second most concerning difficulty overall (with only unexpected seizures being rated more highly). However, when McAuley et al. (2010) posed the same question of clinicians, although the most concerning Epilepsy-related complaint was comparable across patients and clinicians, other responses were dissimilar. Memory difficulties was the twelfth most frequently selected response and did not feature in the top five most concerning difficulties of the disorder overall.

This mis-match between patient and clinician reports of the disorder is relevant to our study, as it reflects the mis-match Epilepsy research has found between subjective and objective memory test scores, and highlights a need to address memory concerns. Use of materials with greater ecological validity and clinical practicality may increase the clinical relevance of memory research, by allowing clinicians to assess this phenomenon more easily, with greater confidence that it is in concordance with patients' real-life experiences, and in parallel with

measuring condition-related factors to ascertain where treatment may best be directed (e.g., memory rehabilitation versus depression/ anxiety management).

1.2 Memory theories

There are several memory theories relevant to this study, which explain the storage and retrieval of declarative memory at different time points after acquisition. The first theory to fractionate memory into temporally-specific storage systems was the Dual-Trace Theory (Atkinson & Shiffrin, 1968). This theory proposed separate short-term and long-term memory stores, which is well-supported by literature demonstrating a disconnection between the two after damage. For example, amnesic patients who could correctly recite a short string of numbers immediately after presentation, but were unable to retain and recall this information over a longer period of time (Baddeley & Warrington, 1970).

Memory consolidation is the process by which information (memory traces) is stabilised after acquisition (Dudai, 2004). Neurologically speaking, the Dual Trace View suggests that memory consolidation occurs after a matter of seconds/ minutes: Newly-acquired information is said to be maintained in a short-term memory store for several seconds or minutes via neuron firing in perception centres of the neocortex. Beyond this time, hippocampal structures bind the information from these perception centres together (e.g., visual, aural, olfactory) to form a single representation of the memory for long-term memory (Hebb, 1949, as cited in Weingartner & Parker, 2014; Squire & Zola-Morgan, 1991).

However, more recent theories suggest that memory consolidation takes place over an extended period of time (e.g., hours, days or weeks), as opposed to the seconds/ minutes suggested by these earlier models (see e.g., Atkinson & Shiffrin, 1968; Baddeley & Warrington, 1970). As introduced in the Journal Paper, there are two key theories of memory consolidation: The Standard Model (see e.g., Alvarez & Squire, 1994; Squire & Alvarez, 1995; Squire & Stark, 2004) and the Multiple Trace Theory (see e.g., Nadel & Moscovitch, 1997; Nadel, Samsonovich, Ryan, & Moscovitch, 2000). Although both of these theories agree that memory consolidation takes place over an extended time period, they propose alternate neural mechanisms underlying this process.

1.2.1 Standard Model of memory

The Standard Model argues that memory traces are initially linked together by hippocampal structures – a temporary, short-term consolidation process, completed within seconds or minutes. However, as long-term consolidation processes begin, this binding and storage is required less and less as the memory trace is increasingly supported by neocortical structures. Over time, the hippocampus therefore becomes redundant as the neocortex can support the permanent memory trace (both in terms of storage and retrieval) (Alvarez & Squire, 1994; Squire & Alvarez, 1995; Squire & Stark, 2004; Squire & Zola-Morgan, 1991).

Squire and Alvarez (1995) highlight cases of temporally graded retrograde amnesia as support for the Standard Model as they indicate that learnt information/ memories are less affected (or not at all affected) by damage to the hippocampus the longer the period of time after initial consolidation (i.e., as the neocortex is now supporting these long-term memories). Older research has supported the Standard Model, similarly highlighting a temporary role for the hippocampus in the memory consolidation process, but suggesting that instead of initially storing the memory traces the hippocampus (1) signals the neocortex to begin its own formation (Wickelgren, 1979), or (2) helps coordinate the initial formation and maintenance of connections between neocortical structures (Teyler & DiScenna, 1986).

1.2.2 Multiple Trace Theory

An alternative view is proposed by Nadel and Moscovitch (1997). They state that the Standard Model is not extensive enough as it cannot account for the very long or very flat retrograde amnesia gradients observed in some amnesia cases. Recent neuroimaging research also suggests that hippocampal activation is equally as activated for recent and remote memories (see e.g., Nadel et al., 2000). This model suggests that the initial process offered by the Standard Model holds true, but that the hippocampus has a more permanent role to play in memory storage and retrieval, particularly for episodic memories (as opposed to the temporary role of the hippocampus and the unitary declarative memory system that the Standard Model proposes).

Episodic memory traces are argued to be re-activated every time a memory is retrieved, creating an associated similar memory trace (note: associated, but not an exact duplicate as traces will have differences dependent on the re-activation context). Over time, successive re-activation leads to the formation of multiple similar traces – trace multiplication. Factual information can be extracted from this collection of traces over time and integrated into semantic memory stores separate to the hippocampus (e.g., the neocortex). Therefore, memories slowly become independent from the initial hippocampal memory trace (a process akin to long-term consolidation). However, spatial and temporal information is still dependent on hippocampal stores and therefore the hippocampus has a permanent role to play in storage and retrieval of episodic memories, despite memory traces being consolidated more sparsely across neurological structures. This creation of multiple traces that overlap and dispersion of storage enables some stability of memories despite hippocampal damage.

Although, Squire and Bayley (2007) highlight that during neuroimaging, or any other testing procedure, the participant will be also be encoding details about the test being carried out (or re-encoding the old memory they are being asked to retrieve in the new context) and so activation of the hippocampus in these studies may be incidental to the processes being studied.

1.3 Temporal lobes involvement in memory consolidation

The temporal lobes are central to the process of memory consolidation (Alvarez & Squire, 1994). More specifically, the literature proposes that key processes are carried out by structures within the medial temporal lobe (MTL) system (see e.g., Gabrieli, Brewer, Desmond & Glover, 1997; Squire, Stark, & Clark, 2004). The MTL system includes the hippocampal formation (cornu ammonis [CA], dentate gyrus, subicular complex), perirhinal, entorhinal and parahippocampal cortices (Squire et al., 2004). The MTL structures are perfectly orchestrated for memory consolidation due to their numerous and extensive neural connections with the neocortex (Squire & Zola-Morgan, 1991).

The use of Magnetic Resonance Imaging (MRI) and voxal-based morphometry techniques have supported the above literature, allowing exploration of MTL activation when completing memory tasks. Research has highlighted an association between damage to MTL systems and reduced memory performance

(see e.g., Alessio et al., 2006; Bernasconi et al., 2003, 2004; Coste et al., 2002), and has proposed that the MTL can be separated in to two distinct neural bases: The parahippocampal cortex in the posterior MTL region for encoding of newly-learnt information and the subiculum in the anterior MTL region for retrieving past memories (Gabrieli, et al., 1997).

1.4 Memory measures

1.4.1 Specificity of traditional 'objective' memory measures

Although the Journal paper highlighted that traditional 'objective' memory measures are potentially not specific enough to detect consolidation difficulties for those with ALF (Piazzini, Canevini, Maggiori, & Canger, 2001), we acknowledge that there were some studies that found memory impairments when using a 30-40 minute delay period (without the need for a longer 'ALF-specific' delay period) (see e.g., Bell, 2006; Helmstaedter, Hauff, & Elger, 1998; Mameniskiene, Jatuzis, Kaubrys & Budrys, 2006). These studies were therefore examined in more detail. Several limitations in their research methodology were noted, for example, Bell (2006) included post-operative patients in their sample and it is argued that surgery is an interacting factor in memory impairment (Téllez-Zenteno, Dhar, Hernandez-Ronquillo, & Wiebe, 2007), which was not considered by the authors. Helmstaedter et al. (1998) and Bell (2006) also failed to sufficiently match clinical and control participants, meaning the clinical sample had a significantly lower mean IQ than controls. Mameniskiene et al. (2006) did not identify the foci of seizure activity, when lateralisation and localisation of seizures have been found to impact on memory performance (see e.g., Hendriks et al., 2004; Jambaqué, Dellatolas, Dulac, Ponsot, & Signoret, 1993). Finally, both Mameniskiene et al. (2006) and Helmstaedter et al. (1998) found that despite finding memory difficulties in their clinical sample at 30 minutes, extended 'ALF-specific' delays showed significant ALF-related deficits in their clinical sample, as compared to control participants. This suggests that the extended delay may be clinically useful to reveal the true extent of patient difficulties. As a result, conclusions suggesting that conventional 30-40 minute delays suffice were not considered credible.

Furthermore, studies that have found significant ALF impairments in clinical samples at extended delays have been noted to show similar limitations (see e.g.,

Bell, Fine, Dow, Seidenberg, & Hermann, 2005; Butler et al., 2007; Martin et al., 1991). Other variations in the research include choice of study material, format of the delayed task (e.g., free versus cued, recall versus recognition), exploration (or not) of the effect of psychosocial variables (e.g., mood) on memory performance, and length of extended delay period (e.g., 24 hours [Martin et al., 1991], 1 week [Helmstaedter et al., 1998], 8 weeks [Blake, Wroe, Breen, & McCarthy, 2000]) (see Butler and Zeman (2008) for a more thorough review).

1.4.2 Ecological validity and multi-modal memory

There are two ways to examine test ecological validity; by assessment of *veridicality* or *verisimilitude* (Franzen, & Wilhelm, 1996). Veridicality refers to the ability of a test to predict performance in everyday life (e.g., as assessed through behavioural observations, self-report). Verisimilitude refers to the ability of a test to place the same demands on an individual as would be expected in everyday life, so that knowledge can be assumed about an individual's ability to perform tasks (Spooner, & Pachana, 2006). Chaytor and Schmitter-Edgecombe (2003) reviewed the literature in this area and concluded that the verisimilitude approach was the most effective way to ascertain a high level of ecological validity, with neuropsychological tests using this method being better predictors of day-to-day cognitive functioning than more traditional test measures.

The difficulty with traditional objective memory measures, is that they are assumed to have veridicality when research examining this fact is minimal (Sbordone, 1996; Spooner, & Pachana, 2006). Traditional objective memory measures were designed with detection and location of neuropathology in mind, as opposed to how they are now used (to predict functional deficits in day-to-day life) (Spooner, & Pachana, 2006).

As memory is suggested to be multi-modal (see e.g. Annett, McLaughlin Cook & Leslie, 1995; Baddeley, Eysenck, & Anderson, 2009; Bigelow, & Poremba, 2014; Gallace, & Spence, 2009), a verisimilitude approach to testing that incorporates all of these modalities in test materials may be best to fully encapsulate an 'accurate' representation of a real-world situation. Furthermore, research into the specific consolidation difficulties experienced with ALF suggests that there are deficits in retrograde memories, particularly of autobiographical information, *across modalities* (Fitzgerald, Mohamed, Ricci, Thayer, & Miller, 2013a).

However, there is some ambivalence about whether this is dependent on the focal hemisphere of the TLE (see e.g., Blake et al., 2000; Helmstaedter et al., 1998; Wilkinson et al., 2012; Fitzgerald, Thayer, Mohamed, & Miller, 2013b). For example, Blake et al. (2000) found that only those with left hemisphere foci showed deficits with verbal information at an extended delay, but other researchers, such as Wilkinson et al. (2012), found that hemisphere specificity was apparent at first recall, but disappeared at long-term follow-up. The variation in materials used to assess ALF makes it hard to fully assess this, as the majority of the literature uses purely verbal memory tests to assess the phenomenon (Fitzgerald et al., 2013a), where multi-modal methods may be best to capture the complex presentation.

Our study using the ALFIE fits well with the above ecological validity research as it 1) assessed test veridicality through examination of correlations with a subjective memory measure (which, although we acknowledge that veridical responding in clinical populations [such as TLE/ Epilepsy sufferers] may be more questionable, would seem appropriate given the healthy population sample used), and 2) used a verisimilitude approach to development of test materials through continued use of real-life televised news broadcasts (see Extended Method). Furthermore, the ALFIE allowed for examination of the variety of deficits across modalities experienced by those with TLE, by embracing a multi-modal approach to assessment.

1.4.3 Subjective memory measures

Subjective memory measures are often used due to assumptions that they are more veridical/ ecologically valid than objective memory measures (Higginson, Arnett, & Voss, 2000). However, as described in the Journal paper, there are several limitations of self-report measures, some of which will be expanded upon in the following section.

If subjective memory measures were assumed to be more veridical than objective memory measures then it might be expected that scores on self-report measures would always indicate greater severity of memory impairment than scores from objective memory tests could ascertain. However, research has not found this to be the case, with Epilepsy patients being seen to both over- and under-estimate

their difficulties (Andelman, Zuckerman-Feldhay, Hoffien, Fried, & Neufeld, 2004; Herrman, 1982).

Hall, Isaac and Harris (2009) argued that this may be due to subjective memory measures being unable to differentiate between memory difficulties and impairments in other cognitive domains. Indeed, scores on subjective memory measures have been found to be predicted by performance on language ability tasks (verbal fluency and vocabulary tests) (Helmstaedter, & Elger, 2000). However, an alternative suggestion is that an individual's ability to accurately assess their own memory becomes difficult in the setting of their memory impairment (Sunderland, Harris, & Baddeley, 1983), meaning use of self-report measures is then problematic. Given that subjective memory measures tap into metacognitive processes (of meta-memory) (Troyer, & Rich, 2002), this is not surprising. Metacognitive knowledge can be split into three types; one type of metacognitive knowledge ('person knowledge') involves the capacity to think about your own cognitive capabilities (Efklides, 2008). When asking individuals to complete self-reports on their memory we are essentially asking them to do this (i.e., asking them to retrieve information from their memory stores, regarding their memory difficulties, so that they can reflect on them). If that individual has a memory problem such as ALF, that causes difficulties with declarative memory consolidation, this would minimise their ability to be able to engage in this process. As a result, that individual would also have difficulty integrating new information (e.g., about current memory failures) into their metacognitive knowledge. This hypothesis would provide a plausible explanation as to why relatives and healthy controls are better able to report memory abilities (see e.g., Helmstaedter et al., 1998).

In conclusion, while subjective memory measures can provide a good overview of patient's experiences, due to the questionable reliability of such measures use of objective memory tests should remain a part of a clinician's assessment of memory functioning.

2. Extended Method

The following sections will outline the study methodology that was reviewed and approved by the University of Lincoln's Research and Ethics Committee (see Appendix H).

2.1 Sample size

In line with the literature, a sufficient sample size for test standardisation was determined as a minimum of 50 participants (when assuming a 95% confidence level) (see e.g., Bridges & Holler, 2007; Crawford & Howell, 1998). Many studies do not provide attrition rates for clinical or control participants. In previous studies an attrition rate of 0% (Mameniskiene et al., 2006) to 16% has been suggested for healthy control participants (Blake et al., 2000). Based on attrition rates in these previous studies it was estimated that in order to achieve our recruitment target we may need to over-recruit with a minimum of 58 participants.

2.2 Recruitment process

The following section will describe the process and rate of recruitment (see also Figure 2).

2.2.1 Identification of participants

Participants were recruited from the general population using a three-pronged approach: 1) contact with members of University of Lincoln staff or undergraduate and postgraduate students enrolled in courses at the University of Lincoln. This was via email advertisement, advertisement on university online groups and through the university's online recruitment site (Sona System), 2) snowballing of participants through the primary author's relatives and peers (whilst still ensuring a diverse range of backgrounds and levels of education), and; 3) contact with local groups, for example; a local church group and several local sales groups.

2.2.2 Initial approach

Potential participants were provided with a Participant Information Sheet (PIS) (see Appendix I) via email or a face-to-face visit (dependent on the identification approach) after enquiring about participating in the study. Interested parties were then provided with an appointment time to meet up and discuss the PIS further and begin Stage 1 of the research. This procedure gave the participants sufficient

time to consider whether they were willing to participate in the study and ask any questions. All potential participants were encouraged to discuss the study with family and friends who may also wish to take part.

Participants were shown a list of the exclusion criteria and asked if any applied to them. Similar to the BMIPB standardisation procedure, they were also asked "Have you ever had an illness or an injury that you have been told may affect your memory?" (Coughlan, Oddy & Crawford, 2007, p. 15) to further establish whether the participant was eligible to take part.

2.2.3 Informed consent

Participants were informed of all information relevant to participation in the study and given the chance to ask questions. It was explained that participation in the study was voluntary and they had the right to withdraw at any time-point (up until 2 weeks after completion of Stage 2) without having to provide a reason. Participants were asked to sign and date two consent forms (see Appendix J); one form was retained for study records, and subsequently stored at the University of Lincoln, and the other form was given to the participant as a record of their consent. Participants would not have been recruited had it been felt that they were unable to give informed consent (please note, this was not an issue during recruitment). Following informed consent, participants embarked on the research process according to the study procedure laid out in the Journal Paper (see also Figure 2).

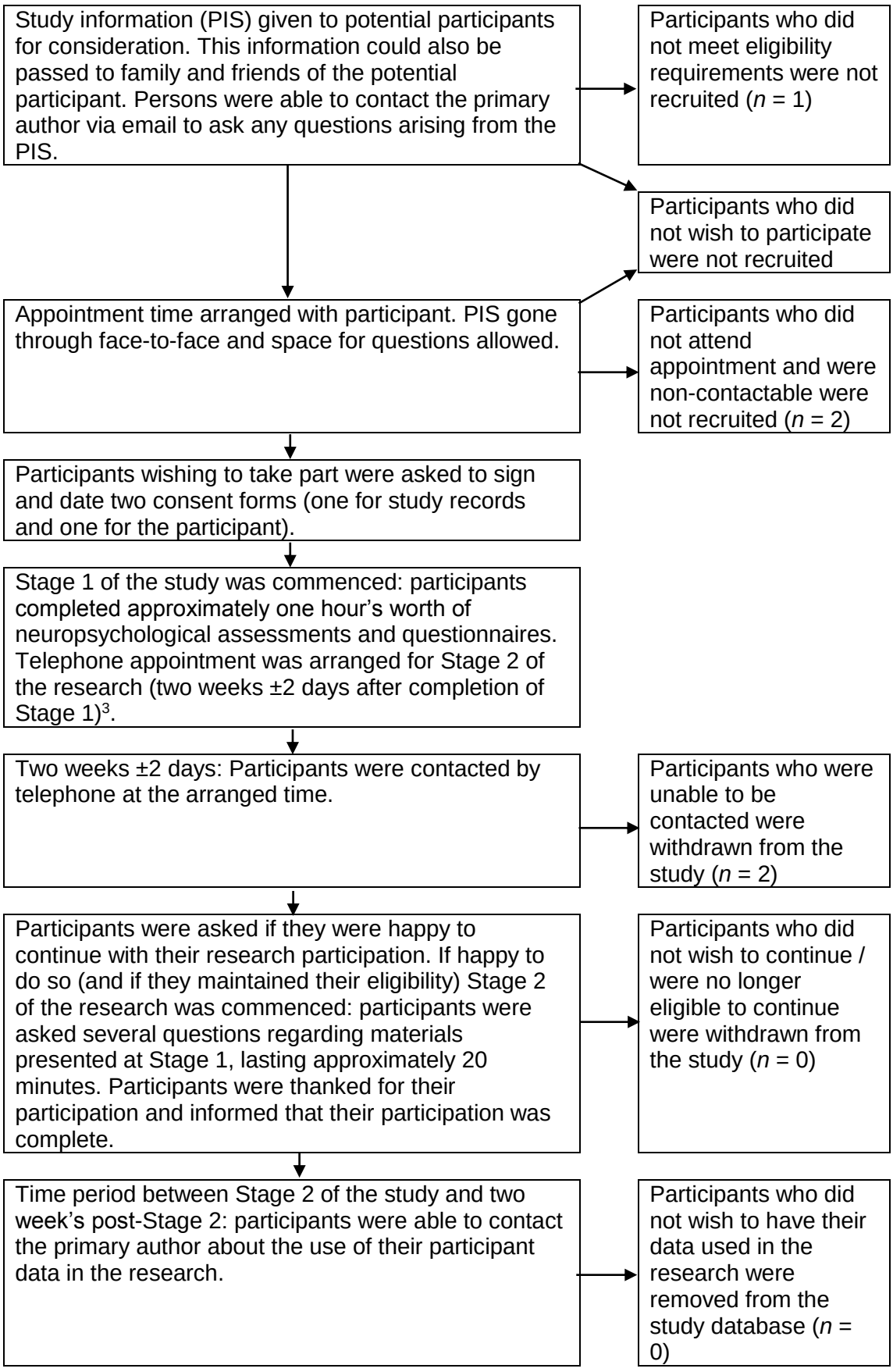
2.2.4 Recruitment rate

In the planning of the study we aimed to recruit a minimum of 50 participants. Based on an estimated recruitment rate of 2-3 participants per week, a recruitment period of 21 weeks was identified. This was extended to 35 weeks due to difficulty recruiting to certain age categories. Hence, recruitment spanned from March 2014 to November 2014.

During the recruitment period several potential participants were not recruited to the study as: 1) after reading the PIS they decided that they no longer wished to take part, 2) they did not attend or cancelled their study appointment and were non-contactable to re-arrange, and; 3) they did not meet eligibility requirements

for the study due to psychotropic medication use (use of anti-depressant medication).

Figure 2. Study regimen



[3] Participants completed a Contact sheet (see Appendix K) to enable telephone contact at the two-week delay.

2.2.5 Instructions for between Stage 1 and Stage 2 of the research

Participants were told that they did not need to do anything between Stage 1 and Stage 2 of the research. There were no instructions given regarding the remembering/ rehearsal of materials and no indication that at Stage 2 participants would be asked to recall and recognise materials again.

McGeoch (1942) stated that it is imprudent to assume that individuals do not have the motivation/ self-instruction to learn, even when given no explicit instructions to do so. Therefore, some participants may have realised that they were engaged in a memory test and may be tested again at Stage 2 and despite having no explicit instruction to remember information until T3, may have self-instructed themselves to continue memorising/ rehearsing stimulus material. However, participants acting in this manner would not be at an advantage over those who did not presume this procedure. Research has long indicated that individuals recall the same or better when they are not primed to remember, compared to those that have been given intentional instructions (see e.g., McDaniel & Mason, 1977; Sparrow, Liu, & Wegner, 2011). Furthermore, there is a small literature base suggesting that rehearsal instructions can decrease memory ability (dependent on age) (Rosner, 1971).

2.3 Demographic variables

2.3.1 Age categories

As research has found a significant relationship between age and ALF in TLE (see e.g., Mameniskiene et al., 2006), as well as between age and long-term forgetting in normal population samples (see e.g., Davis et al., 2003; Mameniskiene et al., 2006; Mary, Schreiner, & Peigneux, 2013) it was felt prudent to consider the impact of this within the present study. In the planning of the study the primary author therefore aimed to recruit a minimum of 50 participants across an 18-75 year age range; the age range used previously for the ALFIE by Corbett (2012). In line with a commonly-used published psychometric test (see Wechsler, 2008) nine age categories were defined to recruit to (see Table 18).

Table 18. Age categories for recruitment

Age range
18.0-19.11
20.0-24.11
25.0-29.11
30.0-34.11
35.0-44.11
45.0-54.11
55.0-64.11
65.0-69.11
70.0-74.11

Assuming recruitment of 50 participants, an average of 5.5 participants was needed within each category. Steps were then taken in the statistical analysis process to examine the effect of such on memory performance and reduce any confounding impact.

2.3.2 Education and predicted full-scale IQ

Several previous studies examining ALF have shown researchers to report either only educational level or IQ of their participants (see e.g., Giovagnoli et al., 1995; Jansari, Davis, McGibbon, Firminger, & Kapur, 2010; Muhlert et al., 2011; Muhlert, Milton, Butler, Kapur, & Zeman, 2010), meaning that it is unsure whether there may be a mis-match between clinical and control samples. For example, Lucchelli and Spinnler (1998) matched their individual clinical case (GB) with two control participants on the basis of years of education (eight years), but neglected to comment on the IQ level of their control sample, despite noting that GB's IQ was above average (120) on the WAIS. It was therefore felt that examining the impact of both of these potentially confounding factors would be fruitful and provide normative data for clinicians to compare against. Steps were therefore taken during statistical analysis to examine the potential impact of these variables on memory performance in the current study.

2.4 Justification of test selection and discussion of their psychometric properties

2.4.1 BMIPB

There are similarities between the BMIPB and well-used equivalents, such as the WMS (see e.g., Wechsler, 2009), for example, the use of a story to assess immediate and delayed verbal memory recall. However, these equivalents are very lengthy to administer and do not provide alternate form versions as the BMIPB does. As two different versions of the ALFIE were going to be analysed, it was felt appropriate to compare the ALFIE to a test that allowed for the same. In an ALF population re-assessment of memory difficulties is common, and repeated assessment using the same form version can lead to content specific practise effects that mean individuals' memory difficulties are not properly captured (Jansari et al., 2010).

The sub-tests of the BMIPB have been found to show good inter-rater reliability, with what Coughlan et al. (2007) coin 'objective measures' (e.g., List Learning) demonstrating absolute inter-rater reliability (r = 1.00) and 'more subjective measures' that are more open to scorer interpretation (e.g., Story Recall) achieving high inter-rater reliability (r = .90). The test shows high test re-test reliability when re-testing on an alternate form (i.e., Version 1 and later Version 2) (Story Recall [immediate and delayed] and List Learning A1-A5: r = .67 – .80, p <.01).

2.4.2 EMQ-28

There are several questionnaire-based measures available to subjectively assess memory. However, many subjective memory measures are criticised for having items that do not relate well to participants completing them, for example questions about public speaking or driving (Emilien, Durlach, Antoniadis, Van der Linden, & Maloteaux, 2004). The EMQ-28 was constructed to relate to everyday experiences participants might have (Sunderland, Harris, & Baddeley, 1984) and was therefore felt to largely avoid this pitfall. Researchers also criticise questionnaires for having varying (and generally poor) validity to identify memory difficulties (see e.g., Hickox, & Sunderland, 1992; Hertzog, & Pearman, 2014). The EMQ was altered from its original 35-item format to a 28-item scale to help combat validity issues and increase the EMQ's validity (Sunderland, Harris, &

Gleave, 1984). Efklides et al. (2002) have more recently studied the EMQ-28's validity and found a strong positive relationship with the Rivermead Behavioural Memory Test, indicating convergent validity with an established ecologically-valid objective memory measure.

2.4.3 WTAR

The WTAR is considerably shorter to administer than its longer Wechsler counterpart, the WAIS. This was useful in the context of our study, which involved a time-costly procedure for participants. Yet, WTAR scores still show good concurrent validity in healthy population samples when correlated with the Verbal IQ Scale from the WAIS (r = .70; Wechsler, 1997). The WTAR has also been shown to have good internal consistency (coefficients for a UK sample ranged from .87-.95; Wechsler, 2001) and good test-retest reliability (coefficients in an American sample ranged from .90-.94; Wechsler, 2001). Furthermore, it has been shown to reliably predict level of educational attainment (Wechsler, 2001) (useful in the context of previous research's limitations, highlighted in section 2.3.2).

In order to allow comparison to future clinical samples, the WTAR was also felt to be the most advantageous measure: Longer measures like the WAIS predict full-scale IQ on the basis of both verbal and performance sub-scales. Performance on some sub-tests of these has been shown to remain stable after brain injury/ disruption (e.g., vocabulary, matrix reasoning), however, other sub-test performance has been shown to deteriorate (e.g., similarities, block design; Green et al., 2008). Reading is highlighted as a key cognitive skill that remains fairly intact despite any neurological insult, hence why measures such as the WTAR and the National Adult Reading Test (NART; Nelson, 1982) are commonly used in research in this area (see e.g., Blake et al., 2000; Muhlert et al., 2011, 2010). The WTAR was chosen over the NART as it is a more recently validated measure and therefore might offer more relevant norms for our sample population.

2.5 Use of BBC East Midlands Today news material

It was checked that the Universities of Lincoln and Nottingham still held an Educational Recording Agency Licence, as had been previously confirmed in Corbett's (2012) study. This meant that we still had permissions to use the clips that Corbett (2012) had previously used, within a research remit (for educational

and non-commercial purposes) (see Educational Recording Agency, 2011). As the aim of the study was to determine whether the ALFIE would be clinically viable as a test (i.e. for commercial use), guidance around continued use of the BBC East Midlands Today news clip was also sought from BBC East Midlands Today and approval gained under the condition that if the test was to enter the public domain it would have to be freely available to clinicians under the BBC license (Kevin Hill, personal communications, March 10, 2014).

2.6 Approach to statistical analysis

The sections below will firstly discuss how data from objective memory measures was evaluated to decide the appropriateness of parametric analysis (section 2.6.1). The assumptions of parametric tests will then be discussed (section 2.6.2), followed by the non-parametric analysis employed for the remainder of the objective memory test data (section 2.6.3). A discussion around use of the kappa statistic to determine inter-rater reliability will follow (section 2.6.4). Finally, the analysis of the subjective memory measure (the EMQ-28) will be examined (section 2.6.5).

2.6.1 Normality of the data

The literature suggests assessing distribution of the data using a range of techniques (see e.g., Kim, 2012, 2013; Pallant, 2010; Tabachnick & Fidell, 2013). The impact of outliers on the dataset was examined by comparing the actual mean with the trimmed mean (obtained by removing the top and bottom 5% of scores). The normality of data distribution was assessed through visual examination of histograms, however it was acknowledged that there is a degree of subjectivity to this (Cribbie, Fiksenbaum, Keselman, & Wilcox, 2012) and so the Kolmogorov-Smirnov test of normality was also employed to provide a more objective measurement. The Kolmogorov-Smirnov test tests the null hypothesis that the data is normally distributed and suggests normality of the data if the *p* value >.05 (Kim, 2012; Pallant, 2010). The distribution of the data was then explored by looking at skewness and kurtosis values. Skewness and kurtosis values indicate a perfectly symmetrical normal distribution if both of these values are zero. A positive skew would indicate that the data was more concentrated towards the lower end of the dependent variable scale (memory performance), with the ‘tail’ of the histogram being longer or ‘fatter’ on the right side of the graph. A positive kurtosis would indicate a higher peak of the data distribution. Negative

skew and negative kurtosis would indicate the converse of these descriptions (Pallant, 2010).

It is recognised that the performance of tests of normality is, however, affected by sample size (Seier, 2002) and interpretation of skewness can be problematic when histogram tail shape may be long on one side of the graph, yet 'fat' on the opposing tail (von Hippel, 2005). Furthermore, the definition of kurtosis varies in the literature and the most widely-accepted definition (as above) is suggested to only be applicable when the skewness value is zero (i.e. when the dataset is perfectly symmetrical) (Balanda & MacGillivray, 1988; Darlington, 1970). Therefore, decisions regarding normality and distribution of the data were felt to be best concluded through close examination of both the visual input, the normality *p* value and skewness and kurtosis values, rather than prioritisation of one method. This is a well-accepted method within the literature (see e.g., Kim, 2013). As described by Kim (2013) a z-test was applied to skewness and kurtosis values, converting them into z scores using the formulae below:

$Z_{skew} = Skew/SE_{skew}$

$Z_{kurtosis} = Kurtosis/SE_{kurtosis}$

The null hypothesis would be rejected if the absolute z score value >1.96, indicating non-normal distribution of the data. The results of these assessments are provided in Table 19.

For all tests, assessment of data distribution was carried out for the earliest time-point possible (i.e. T1 or T2 dependent on sub-test). Assessment was not carried out at all time-points as it was unclear how the data should behave at later time-points. Comparison of the actual mean with the trimmed mean indicated that data was not notably affected by outliers. The Kolmogorov-Smirnov statistic indicated that only the BMIPB story recall was normally distributed at T1. However, visual examination and, to a great extent, skewness and kurtosis values, also indicated that the ALFIE narrative and visual recognition tasks were normally distributed at T2. Across the dataset, negative skewness could be observed for all sub-tests apart from the ALFIE story recall, where data showed a positive skew. Kurtosis

values were variable, with patterns of both peaked and flattened distributions (see Table 19).

Table 19. BMIPB and ALFIE normality data

	BMIPB				ALFIE		
	Story recall (T1)	List			Story recall (T1)	Narrative recognition (T2)	Visual recognition (T2)
		Recall (T1)	Word recognition (T2)	List recognition (T2)			
Mean	29.02	11.67	26.96	27.65	11.27	11.56	9.88
5% trimmed mean	29.25	11.92	27.09	28.08	11.15	11.63	9.91
Visual examination evaluation[a]	Y	N	N	N	N	Y	Y
Skewness z score[b]	-1.03	-4.46	-3.19	-7.29	1.01	-2.10	-.80
Kurtosis z score[b]	-.53	6.09	1.30	11.76	-1.51	1.05	-.22
Kolmogorov D	.11	.12	.21	.22	.15	.19	.21
- Smirnov *p*	.09	.05	.01	.01	.01	.01	.01

[a]*Where Y indicates normal distribution of the data from visual examination and N indicates non-normal distribution of the data.* [b]*Where non-normal distribution is indicated by an absolute value >1.96; BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); T1 = Immediate recall; T2 = 40-minute delay.*

2.6.2 Parametric analysis for objective memory test data

Parametric statistical tests are known for possessing greater power to detect an effect than non-parametric equivalents, meaning a decreased risk of Type II error (Langdridge, 2004; Pallant, 2010). However, the power of these tests is compromised if assumptions of the tests are violated, for example; the assumption that data is normally distributed (Langdridge, 2004; Pallant, 2010). As a result of the above assessment of normality, only the BMIPB story recall task and the ALFIE recognition tasks were therefore analysed using parametric statistics.

2.6.2.1 Approach to conducting mixed ANOVAs

Before conducting mixed ANOVAs single case representatives were excluded to allow testing for interactions and reduce the chance of erroneous results (Type I and Type II error). This meant that two participants were removed from the analyses for V1 (the one participant who was in the age group 70-74.11 and the one participant who had ≥22 years of education) and four participants were removed from the analyses for V2 (the one participant in age group 30-34.11, age group 35.0-44.11, age group 65.0-69.11 and ≥22 years of education, respectively). Several test assumptions were also examined (above and beyond ensuring normal distribution of the data) as recommended by Langdridge (2004). These are outlined in the following sections.

2.6.2.1.1 Design

Data for a mixed ANOVA should contain both repeated measures and independent measures. The within-participant (repeated measures) independent variable should consist of at least two categorical related groups and the between-participant independent variables should consist of at least two categorical independent groups (Langdridge, 2004). In the current study the variables were therefore ‘categorised’ as shown in Table 20, in order to conduct 2x5x9 three-way mixed ANOVAs for objective memory test scores (dependent variable).

Table 20. Independent and dependent variables for mixed ANOVAs

Independent variable type	Factor	Conditions
Between-participant	Gender	Male
		Female
	Years of education	5-11 (Primary education)
		12-16 (GCSEs or equivalent)
		17-18 (A-Levels or equivalent)
		19-21 (Degree-level or equivalent)
		≥22 (Above degree-level or equivalent)
	Age	18.0-19.11
		20.0-24.11
		25.0-29.11
		30.0-34.11
		35.0-44.11
		45.0-54.11
		55.0-64.11
		65.0-69.11
		70.0-74.11
Within-participant	Time	T1
		T2
		T3

2.6.2.1.2 Outliers

There should be no significant outliers in any within-participant or between-participant independent variable groups. Outliers can increase the estimate of sample variance, which can decrease the value of the *F* statistic and increase the chance of Type II error. Trimmed means indicated no adverse effects of outliers (See Table 19).

2.6.2.1.3 Homogeneity of inter-correlations/ Equality of covariance matrices

Box's *M* statistic tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups (Pallant, 2010). Tabachnik & Fidell (2013) suggest that this test is highly sensitive and so state to only assume significance of the *M* statistic when $p < .001$. Non-significant test results for covariance matrices across all objective memory scores at all time-points

indicated that this assumption was not violated (BMIPB recall: M = 3.32 – 71.95, p ≥.02; BMIPB recognition: M = 1.60 – 23.60, p ≥.01; ALFIE recall: M = 1.40 – 41.63, p ≥.14; ALFIE recognition: M = 1.14 – 26.31, p ≥.08).

2.6.2.1.4 Homogeneity of variance

Levene's test assesses the null hypothesis that the error variance of the dependent variables is comparable across groups. This diagnostic test was therefore conducted for the dependent variable with each between-participants variable (age, gender, years of education). The *F*-statistic is suggested to be fairly robust against inequality of variance if sample sizes are roughly equal. However, when the sample variances are very different to one another there is a greater chance of incorrectly rejecting the null hypothesis (Type I error) (Keppel & Wickens, 2004; Tabachnik & Fidell, 2013). Keppel & Wickens (2004) suggested that there is no consistent point at which unequal sample sizes make heterogeneity of variance an issue. However, various researchers suggests that ANOVAs are robust to heterogeneity of variance so long as the largest variance is not more than nine or ten times the smallest variance (the F-ratio; Keppel & Wickens, 2004; Tabachnik & Fidell, 2013). This ratio rule (of 1:10) was therefore upheld when examining Levene's test outputs.

Where homo-scedasticity could not be assumed Keppel & Wickens (2004) suggest adopting a smaller alpha level (α = .025). This recommendation was therefore complied with when heterogeneity of variance was observed, in order that parametric statistics could be employed. Although it is acknowledged that this does reduce the power of the *F*-statistic it was felt that this would still allow for use of a more robust statistical measure.

Applying the conditions stated above, all error variances were found to be equal for all objective memory scores at all time-points, aside from for the ALFIE V2 story. With the ALFIE V2 story the homogeneity of variance assumption was violated for the between-groups variable of years of education for objective memory scores at T3 (F(3,17) = 9.77, p =.01). The ratio of the variance in this case was 1:49 (see Appendix L). A .025 alpha level was therefore used when analysing the ANOVA output for this dataset.

2.6.2.1.5 Sphericity

Mauchly's test of sphericity examines the equality of variances of the differences between the within-participant independent variables groups for each individual between-participant factor groups. Mauchly's test was only significant for the BMIPB story recall, indicating that the assumption of sphericity was violated ($X^2(2) = 9.64$, $p = .01$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = .70$).

2.6.3 Non-parametric analysis for objective memory test data

There is much discussion about which way to approach non-parametric data for analysis (see e.g., Hoboken, Wolfe, & Chicken, 2014; Kim, 2014a, 2014b; Wasserman, 2006). Non-parametric statistics could be employed, however, non-parametric statistics are less powerful than their parametric equivalents. The use of parametric bootstrapping is highlighted in the literature as an effective method, particularly with small samples sizes, such as in the current study (Cribbie et al., 2012; Krishnamoorthy, Lu, & Mathew, 2007) and this option was therefore explored.

The bootstrap methodology was first discussed by Efron (1979) and is a resampling method whereby you treat your sample as a proxy for your population and resample from this 'population' multiple times to simulate repeated sampling of the actual population (as resampling from the actual population may be unfeasible). As a result, you can produce an estimate of the sample distribution (as opposed to the *population* distribution, as one assumes this when assuming your sample is an adequate model for your population), meaning you can then report estimated confidence intervals based on these multiple resamples (bootstrapped confidence intervals) (Efron, 1979; Krishnamoorthy et al., 2007). The main assumption of this method is that your original sample is a good representative of the actual population. Cribbie et al. (2012) suggest that trimmed parametric bootstrapping (i.e. the removal of outliers via trimmed means) is suggested to be the most robust method of reducing Type I error and maintaining good statistical power. As the present study did not note any adverse effect of outliers (see section 2.7.1) it was felt that this additional step was supernumerary.

Where possible, non-parametric data was therefore examined using the same parametric tests as parametric data, but through utilisation of the parametric

bootstrap method. Data was trimmed to remove single case representatives similar to for parametric analysis. Test outputs were reported with 95% bootstrapped confidence intervals.

For all other non-parametric analyses, rather than violate the assumptions of parametric tests, non-parametric equivalents were used. For example, the Kruskal-Wallis test was used to assess any significant differences in memory performance (on the BMIPB Word list recall and recognition tasks, and the ALFIE Story recall) between the demographic variables of gender, age and years of education. However, parametric tests were also used on this data to allow comparison of non-parametric test outcomes to parametric test outcomes. This meant that whilst we could acknowledge that assumptions for parametric testing had not been met, where outcomes were the same, the parametric test outcome could be reported in the Journal Paper as it afforded more precision and power.

2.6.4 Inter-rater reliability

The Kappa statistic, developed by Cohen (1960), is commonly used in the literature to assess inter-rater reliability (McHugh, 2012). Assessment of inter-rater reliability is important as it is likely that there will be a variety of professionals scoring data in a clinical setting, who may interpret responses differently. The kappa statistic therefore looks at the amount of consistency, or agreement, between individuals. It also controls for random agreement factor (the hypothesis that scorers will sometimes be congruent as a result of guesses, rather than fully knowing the correct way to score), something that other inter-rater reliability analyses do not (e.g., percent agreement) (McHugh, 2012). The kappa statistic can range from -1 to +1, although a value below zero is improbable. A score of zero represents the amount of agreement that would be expected by chance and a score of +1 indicates perfect congruence between scorers. Any score <1 would therefore not only indicate the amount of agreement between raters, but also the amount of *disagreement*. A kappa value of; ≤0 indicates no agreement, .01 – .20 none to slight, .21 – .40 fair, .41 – .60 moderate, .61 – .80 substantial, and; .81 – 1.00 as almost perfect agreement (Cohen, 1960; Landis & Koch, 1977). It is generally accepted that values <.60 in healthcare and clinical settings would suggest inadequate inter-rater reliability of the measure (McHugh, 2012). As McHugh (2012) suggests that kappa statistics are often misinterpreted a percent agreement statistic has also been reported.

2.6.5 Analysis of the subjective memory measure

Data distribution of the EMQ-28 scores was assessed using the same methods as discussed in section 2.6.1. Comparability of the means and the 5% trimmed means indicated no adverse effect of outliers on the data. Both the EMQ-28 total score and the EMQ-28 score for Q29 were found to be non-normally distributed (see Table 21). Kolmogorov-Smirnov tests were significant for both datasets at the .01 level and both datasets showed a positive skew (indicating better memory performance, which would be expected in a normal population sample). As with the distribution of objective memory test scores, this indicated whether parametric or non-parametric statistical analysis should be employed (for actual tests used see Journal Paper).

Table 21. EMQ-28 normality data

		EMQ-28	EMQ-28 Q29
Mean		42.56	1.85
5% trimmed mean		40.77	1.70
Visual examination evaluation[a]		Y	N
Skewness z score[b]		3.40	3.57
Kurtosis z score[b]		2.23	.42
Kolmogorov-Smirnov	D	.18	.26
	p	.01	.01

[a]*Where Y indicates normal distribution of the data from visual examination and N indicates non-normal distribution of the data.* [b]*Where non-normal distribution is indicated by an absolute value >1.96.*

3. Extended Results

3.1 Discussion of issues surrounding exclusion on the basis of IQ

One participant who completed the current study was predicted to have a full-scale IQ of 62 (as assessed using the WTAR). As the normative range for IQ is accepted to be between the ranges of 70-130 (as assessed by two standard deviations from the mean [of 100]), this participant was deemed to be outside of a normal population sample. They were therefore removed from the dataset and excluded from analysis as it was felt that their data may have skewed the results.

PFSIQ was not listed as a pre-defined exclusion criterion as participants' PFSIQ was unknown at the outset the research. However, participants PFSIQ would have been known by Stage 2 of the research, due to the WTAR being completed during Stage 1, between T1 and T2 recall/ recognition. This therefore presented an ethical dilemma for us: Inform participants at Stage 2 of the research that, based on their PFSIQ, they were no longer eligible to take part in the study, or allow participants to continue with the study until completion, despite knowing that their time investment was valueless as their data could not be utilised and would be removed from the dataset.

As participants were informed at the outset that they would not be receiving any results/ feedback on their performance on tasks, due to the large number of individuals needing to be recruited and the limited time available on the part of the primary author, it was felt that later feeding back results regarding PFSIQ would be unethical. Participants would have been unprepared for this information and would not have been given a choice about whether they wished to receive this information. Furthermore, the WTAR has a focus on verbal IQ and the accuracy of this prediction when generalising to PFSIQ cannot be assumed, despite the sound psychometric properties of the test (see section 2.4).

3.2 Approach to percentile creation

The following sections will outline the decision-making process of how to transform raw scores from the ALFIE and from the extended sections of the BMIPB (section 3.2.1), and then the definition of the chosen transformation method (section 3.2.2).

3.2.1 Choice of transforming to percentiles

Although many researchers show a preference to standardised scores over percentile ranks (see e.g., Bowman, 2002; Crawford, 2004; Lezak, Howieson, Loring, Hannay, & Fischer, 2004), percentiles were chosen to represent individual performance on the ALFIE. Standardised scores were inappropriate due to some data being non-normally distributed. It also allowed direct comparability with the BMIPB, which converts raw scores into percentile ranks. Furthermore, it has been highlighted that the transparency of meaning of percentile ranks lends itself well to communicating results to lay-persons (Crawford, Garthwaite, & Slick, 2009; Lezak et al., 2004), which was deemed important for when testing levels of ALF in epileptic patients in a clinical setting.

3.2.2 Percentile definition

Researchers are often unclear as to the definition of a percentage they have chosen when transforming raw scores (Crawford et al., 2009). There are three possible definitions:

> definition A -the percentage of scores that fall below the score of interest, B - the percentage of scores that fall at or below the score of interest), and C -the percentage of scores that fall below the score of interest, where half of those obtaining the score of interest are included in the percentage (Crawford et al., 2009, p. 5).

It is apparent that transparency of the percentile definition used is key, as the differing definitions defined above can produce varying percentile ranks from the same raw score, particularly for small sample sizes (Crawford et al., 2009). We are therefore explicit with regards the definition we have used, as is outlined in the following paragraphs.

The best method of transforming raw scores into percentile ranks is suggested to be Definition C (Crawford et al., 2009). As described by Ley (1972, as cited in Crawford et al., 2009; Stockburger, 2013), percentiles for raw scores would then be calculated as follows:

$$\text{percentile rank} = \left(\frac{m + .5k}{N}\right)100,$$

where *m* is the number of scores in the sample that are below a given score, *k* is the number of scores in the sample that are of the same value as the given score, and *N* is the total number of scores in the sample. This was therefore the definition our study aimed to adopt. However, unfortunately it was felt that this was not the best method of transformation for the data in the current study as calculating via Definition C gave very specific percentile ranks and it was felt that for such a small population sample this level of precision would not be accurate. Furthermore, it would not allow direct comparison to published objective memory measures that tend to utilise landmark percentiles (e.g., the BMIPB). Definition A was therefore utilised, with landmark percentiles chosen in line with the BMIPB.

4. Extended Discussion

4.1 Impact of potentially confounding variables on memory performance

In the Journal paper factors such as diet, physical exercise, stress levels and mood were cited as potentially confounding variables on memory performance (see e.g., Bell & Giovagnoli, 2007; Butler & Zeman, 2008; Chepenik, Cornew & Farah, 2007; Small, 2002). Although exclusion criteria and counterbalancing of the sample were used to minimise the impact of these variables, it was felt important to orientate the reader to the potential impact of these factors and they will therefore be expanded on in the following sections.

4.1.1 Impact of diet

Studies suggest that poor glucose regulation is related to poor cognitive performance (including memory abilities) and smaller hippocampal volumes (see e.g., Convit, Wolf, Tarshish, & de Leon, 2003; Kaplan, Greenwood, Winocur, & Wolever, 2000; Messier, Desrochers, & Gagnon, 1999), a brain area highlighted to be key for memory consolidation (see Extended section 1.3). The consumption of dietary macronutrients is said to improve these abilities (see e.g., Jones, Sünram-Lea, & Wesnes, 2012; Kaplan et al., 2000; Meikle, Riby, & Stollery, 2004). It has long been proposed that the cognitive function most affected by the consumption or deficiency of these nutrients is declarative memory (Manning, Hall, & Gold, 1990; Manning, Parsons, Cotter, & Gold, 1997; Coccoz, Sandoval, Stehberg, & Delorenzi, 2013). Several studies have found that the impact of these macronutrients shows temporal variation, with significant positive associations with memory performance shown at 15 minute and 60 minute delated recalls, as well as increased rates of forgetting (dependent on the macronutrient examined) (see e.g., Jones et al., 2012; Kaplan, Greenwood, Winocur, & Wolever, 2001). These findings may be important to consider when examining accelerated long-term forgetting.

4.1.2 Impact of physical exercise

Physical exercise is proposed to increase cognitive functioning in a range of areas, including memory performance (Erickson et al., 2011; Stroth, Hille, Spitzer, & Reinhardt, 2009; Winter et al., 2007), and also increase (or preserve) hippocampal and MTL volume (Erickson et al., 2009, 2011; Honea et al., 2009). However, it is postulated that it is the *intensity* of the exercise and the

engagement in a *variety* of exercise activities, not the duration of the exercise, which is key (Angevaren et al., 2007; Podewils et al., 2005). Although the literature does highlight that this association may be mediated by other factors, for example cardiovascular factors or neurotrophin levels (see e.g., Kirk-Sanchez, & McGough, 2014).

4.1.3 Impact of stress

The impact of stress levels on memory performance varies in the literature, dependent on when stress is induced. For example, increased stress levels close to the point of information acquisition appear to enhance memory performance (see e.g., Cahill, & Alkire, 2003; Cahill, Gorski, & Le, 2003), however, increased stress levels at or near the point of retrieval are associated with increased long-term forgetting (see e.g., Roozendaal, 2002; Trammell, & Clore, 2013). This association with delayed recall remains despite variations on mode of stimuli to be remembered (verbal/ visual), emotional salience of stimuli, and opportunities for rehearsal (Trammell, & Clore, 2013). The higher cortisol levels are, the greater the impact on memory performance (Buchanan, Tranel, & Adolphs, 2006; Tollenaar, Elzinga, Spinhoven, & Everaerd, 2008). Declarative memory performance has been found to be significantly impaired by stressful conditions, with non-declarative memory being unaffected (Lupien et al., 1997). However, it is recognised that the study of the impact of stress on memory performance is limited somewhat by the ecological validity of such studies – for example, inducing stress by immersion of a limb in icy water (Trammell, & Clore, 2013) or through epinephrine administration (Cahill & Alkire, 2003).

4.1.4 Impact of mood

Studies have long reported that mood (for example, depression or anxiety) is negatively associated with memory performance (see e.g., Bornstein et al., 1991; Kizilbash, Vanderploeg, & Curtiss, 2002; Fossati et al., 2004), with the presence of co-morbid mood disorders further exacerbating difficulties with acquisition and retrieval (Kizilbash et al., 2002). However, it is important to note here that the TLE literature generally suggests that ALF is not associated with mood (Blake et al., 2000; Butler et al., 2009; Howard et al., 2010; Muhlert et al., 2011; Wilkinson et al., 2012).

4.2 Clinical Implications

This study proposes that the ALFIE test may be a clinically viable measure: It was designed using the verisimilitude approach, showed good validity and reliability (e.g., correlated well with a subjective memory measure (showing veridicality [with at least a healthy population sample], showed convergence with a published objective memory measure), and showed promise as a measure that would be practical to implement for Epilepsy clinics (due to favourably low attrition rates).

Use of a telephone method for two-week delayed assessment might be a clinically viable solution for specialist services assessing ALF over large geographical areas, as it is less costly, more practical and more easily organised than asking patients to come back into clinic within such an exact time-scale. However, it is acknowledged that this research was conducted with a non-clinical sample and so qualification of this statement may be needed now that standardised norms have been created. For example, some factors might be expected to lower attrition in a clinical sample (e.g., potentially more intrinsic interest/ motivation), but there is the potential to increase attrition too (e.g., might forget prearranged phone appointment or intrinsic importance may actually potentiate avoidance [e.g., possibility of confirming fears about memory]).

4.3 Directions for future research

Several avenues of future research were highlighted as a result of our study. The following sections aim to extend on the description of some of these that were provided in the Journal Paper.

4.3.1 Comparison of two-week delayed assessment methods

Comparison of face-to-face and telephone methods of delayed follow-up would be useful. To our knowledge no published studies have assessed this and it is unclear what impact remote assessment has on testing. For example, as immediate and 40-minute delayed recall are conducted in more clinical (and arguably less ecologically valid) settings than the two-week delay (where influencing external factors cannot be as well controlled for).

4.3.2 Use of a 0-1-2 scoring system

Participants were sometimes observed to recall ALFIE story information, yet not provide enough detail to score a point for that information unit. For example, a

participant who completed ALFIE Story V1 recalled that “The council overturned the fine.” This was a combination of two information units: 1) that “The city council has started an enquiry,” and 2) that “the fine was immediately overturned.” In this example, the participant would receive one point for the recall of the fine being over-turned, but zero points for council involvement as they failed to mention an investigation. Similarly, for the information unit “Two years ago Sarah was awarded an MBE” (from ALFIE Story V2) one participant recalled that “[she was] already awarded an MBE.” Despite remembering the specific award (MBE) from the story, the participant obtained zero points as they were unable to also recall that it had been two years ago. We therefore wondered whether a 0-1-2 scoring system may be beneficial in providing a more detailed account of forgetting for clinicians. This type of scoring system would also be in line with published objective memory measures, such as the BMIPB.

A 0-1-2 pilot scoring system was therefore devised for the ALFIE Story V2 (see Appendix M) as an alternative to Corbett’s (2012) 0-1 marking criteria. This would allow a maximum of 56 points to be awarded for recall at T1, T2 and T3 (As opposed to Corbett’s [2012] maximum of 28 points). The structure of the 0-1-2 scoring system was in line with a published objective memory measure, where: 2 = a correctly recalled/ paraphrased information unit, 1 = a vaguely or partially-recalled information unit, and 0 = an incorrect/ additional information unit (Coughlan et al., 2007). For example, for the first information unit: ‘News from Sarah,’ an acceptable two-point answer would be ‘Sarah,’ an acceptable one-point answer would be the whole name replaced by ‘a woman’/ ‘a lady,’ and a zero-point answer would be the whole name replaced by ‘a person’/ wrong name given. 10% of participants from the current study (*n* = 5) were then scored according to this proposed new criteria to examine whether there was any relationship between their ‘0-1’ and ‘0-1-2’ raw scores at T1, T2 and T3, and thus determine whether a 0-1-2 scoring system may be clinically useful.

Information that scored one-point according to the 0-1 scoring system was assumed to meet the criteria for two-points according to the 0-1-2 scoring system. All 0-1 raw scores were therefore doubled before statistical testing was carried out, to enable a more accurate comparison to the new 0-1-2 scoring system. Descriptive ‘difference scores’ between these two totals can be seen in Table 22. A ‘difference score’ was calculated as follows:

Difference score = *b* – (*a**2),

where *a* is the total raw score from the 0-1 scoring system, and *b* is the total raw score from the 0-1-2 scoring system.

Table 22. Impact of a 0-1-2 scoring system on ALFIE Story V2 raw scores

Ppt no.		Scoring system raw score		Difference score[a]
		0-1	0-1-2	
8	T1	5	11	1
	T2	2	7	3
	T3	1	4	2
10	T1	4	9	1
	T2	6	13	1
	T3	2	3	-1
21	T1	11	20	-2
	T2	11	21	-1
	T3	6	12	0
35	T1	6	12	0
	T2	5	10	0
	T3	1	4	2
46	T1	7	14	0
	T2	7	12	-2
	T3	6	11	-1

Ppt no. = Participant number; ALFIE = Accelerated Long-term Forgetting In Epilepsy test (Corbett, 2012); V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; [a]*Difference score = the difference between a 0-1-2 raw score and a doubled 0-1 raw score.*

Bootstrapped Pearson's correlation coefficients found significant relationships between the doubled 0-1 raw scores and the 0-1-2 raw scores at all time-points (T1: $r(5) = .99$, 95% CI [.98 – .99], $p = .01$; T2: $r(5) = .97$, 95% CI [.62 – .99], $p = .01$; T3: $r(5) = .97$, 95% CI [-1.00 – 1.00], $p = .01$). It is therefore suggested that a 0-1-2 scoring system would not add significant value to clinicians over and above the 0-1 scoring system already proposed by Corbett (2012).

4.3.3 Development with clinical populations experiencing ALF

Future examination with clinical samples would allow firmer conclusions regarding the lower attrition rates seen in this study (see section 4.2). Furthermore, as the ALFIE is suggested to be a more ecologically valid measure than published memory measures, potential future use of the ALFIE clinically may help to alleviate the discrepancies researchers have found between clinician and patient concerns regarding memory (see e.g., McAuley et al., 2010), as memory difficulties (such as ALF) may be picked up through testing, making it a more salient concern for clinicians.

It would be useful to extend the investigation of the use of memory measures with greater ecological validity to other clinical populations. The set of norms provided by our study could be used/ developed further in this vein. Previous research has indicated that ALF is experienced by those with severe closed-head injuries (particularly when the foci of damage is to the temporal lobe region) (Carlesimo, Sabbadini, Loasses, & Caltagirone, 1997), those with post-traumatic amnesia (Levin, High Jr., & Eisenberg, 1988), in stroke populations (Gold, & Trauner, 2014; Sicong, Miller, Piguet, & Hornberger, 2014), patients with confusional states (such as post-electroconvulsive therapy, delirium) (Lewis, & Kopelman, 1998), semantic dementia (Tu, Mioshi, Savage, Hodges, & Hornberger, 2013) and in some older adult samples (although this is somewhat debated in the literature) (see e.g., Salthouse, 1991).

4.3.4 Replicability 'within-language'

The ALFIE may not be utilisable outside of an English-speaking Western demographic, as we would need to match for language differences (e.g., would need to balance familiarity, complexity and meaning across languages) to be able to then detect differences that are more about culture than linguistic. Furthermore, research examining cross-cultural differences has found cross-cultural gender differences in memory recall meaning that assuming we could match for language differences, findings related to gender differences may not be replicable: females (specifically mothers) are found to provide more detailed and emotional elaborate memory recalls in Western cultures than in other cultures, for example Korea (Mullen & Yi, 1995) or China (Wang, 2006). As a result, determining replicability 'within language' through examination of cultural differences within English-speaking demographics was felt to be a viable future direction.

4.4 Personal reflections on the undertaking of this research

Despite reflexivity being argued to add credibility and rigour to research (De Souza, 2004), personal reflections are not common-place in quantitative research and are sometimes contraindicated, for example due to the dichotomy it creates between the positivist epistemological stance of the researchers and the more constructivist, personal accounts then provided (see e.g., Millen, 1997; Ryan & Golden, 2006). However, it was felt that providing space for reflections would add a depth of understanding and context to the research project (as suggested by e.g., Ryan & Golden, 2006; Walker, Read, & Priest, 2013). Furthermore, as Walker et al. (2013) state, completing a doctoral piece of research is not just about the academic work being completed, but a process of developing one's self and gaining valuable transferrable skills. The following sections will therefore elaborate on my personal motivations for wanting to conduct research in this area, followed by the challenges I faced throughout the project and the resources I drew upon to overcome these difficulties.

4.4.1 Personal motivations

In planning this research, I wanted to focus on an area of clinical psychology that interested me and was relevant to my later career pursuits. By drawing on my previous experience (as an Assistant Neuropsychologist) and my interests I felt that I would be able to make a strong contribution to the research area. My research tutors and I had several discussions about fields of study and it became clear that developing the ALFIE further might be useful clinically, due to the gap in published memory tests available to measure ALF, as well as provide support for more recent psychological theories on memory consolidation.

4.4.2 Challenges faced during the research

Research in this area required me to do a substantial amount of reading on the topic of ALF and TLE, as this was not a field I was familiar with. Getting to grips with the extensive literature base, the technical language in the literature and understanding the intricacies of the neurology of the disease brought a challenge that I had not anticipated. Later, the complexities of data analysis also required extensive reading in order to be able to make informed decisions about, for example; assessment of test assumptions, choice of statistical test.

Recruitment and data collection were time and resource intensive, and tested my organisational abilities. It was tiring travelling to recruit participants and frustrating when, despite the amount of time I had invested advertising and recruiting to the study, I was not going to reach my recruitment target by the pre-planned recruitment deadline. Attempting to score, input and analyse the large amount of data felt overwhelming at times.

4.4.3 Resources employed to overcome the challenges faced during the research

Conducting this research has taken me away from my role as a partner, family member and friend. Paradoxically, I would not have been able to recruit as many participants as I did without the support of these people who I was taken away from. It was heartening, the number of people who were willing to invest time in my research (in particular members of the local community who had no connection with me). I have learnt a lot about myself during this process; gaining an understanding of my methods of coping, my ability to reflect on the value of the research during stressful times and my ability to be able to structure and fractionate tasks to become more efficient. I feel it has taken great personal strength to persevere with this research to completion. However, I also acknowledge that the support I received was invaluable in achieving this. It has been personally important for me to reflect on the positive impact and value of these supportive relationships, across all aspects my life.

This research project has been anxiety-provoking, particularly as the submission deadline drew closer, yet also exciting as I could see my ideas coming together and see the fruits of my labour. I enjoyed having discussions about my research, interpreting the findings and also realising the knowledge I had gained on this journey. Overall, this research project has been a tiring, yet rewarding experience, and one which has helped me develop useful skills for clinical practise and everyday life.

Extended References

Alessio, A., Bonilha, L., Rorden, C., Kobayashi, E., Min, L. L., Damasceno, B. P., & Cendes, F. (2006). Memory and language impairments and their relationships to hippocampal and perirhinal cortex damage in patients with medial temporal lobe epilepsy. *Epilepsy Behav, 8,* 593-600.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences: Neurobiology, 91,* 7041-7045.

Andelman, F., Zuckerman-Feldhay, E., Hoffien, D., Fried, I., & Neufeld, M. Y. (2004). Lateralisation of deficit in self-awareness of memory in patients with intractable epilepsy. *Epilepsia, 45(7),* 826-833.

Angevaren, M., Vanhees, L., Wendel-Vos, W., Verhaar, H. J., Aufdemkampe, G., Aleman, A., & Verschuren, W. M. (2007). Intensity, but not duration, of physical activities is related to cognitive function. *Eur J Cardiovasc Prev Rehabil,14(6),* 825-830.

Annett, J., McLaughlin Cook, N., & Leslie, J. (1995). Interference with olfactory memory by visual and verbal tasks. *Perceptual and Motor Skills, 80,* 1307-1317.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. II, pp. 89-195). New York, USA: Academic Press.

Baddeley, A., Eysenck, M.W., & Anderson, M.C. (2009). *Memory.* London, UK: Psychology Press.

Baddeley, A., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short- term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176-189.

Balanda, K. P., & MacGillivray, H. L. (1988). Kurtosis: A critical review. *The American Statistician, 42(2)*, 111-119.

Bell, B. D. (2006). WMS-III logical memory performance after a two-week delay in temporal lobe epilepsy and control groups. *Journal of Clinical & Experimental Neuropsychology, 28,* 1435-1443.

Bell, B. D., Fine, J., Dow, C., Seidenberg, M., & Hermann, B. P. (2005). Temporal lobe epilepsy and the selective reminding test: The conventional 30-minute delay suffices. *Psychological Assessment, 17(1),* 455-476.

Bell, B. D., & Giovagnoli, A. R. (2007). Recent innovative studies of memory in temporal lobe epilepsy. *Neuropsychology Review, 17,* 455-476.

Bernasconi, N., Bernasconi, A., Caramanos, Z., Antel, S. B., Andermann, F., & Arnold, D. L. (2003). Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region*. Brain, 126,* 462-469.

Bernasconi, N., Duchesne, S., Janke, A., Lerch, J., Collins, D. L., & Bernasconi, A. (2004). Whole-brain voxel-based statistical analysis of gray matter and white matter in temporal lobe epilepsy. *NeuroImage, 23,* 717-23.

Bigelow, J., & Poremba, A. (2014). Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. *PLoS ONE, 9(2),* e89914. DOI: 10.1371/journal.pone.0089914

Blake, R. V., Wroe, S. J., Breen, E. K., & McCarthy, R. A. (2000). Accelerated forgetting in patients with Epilepsy. *Brain, 123,* 472-483.

Bornstein, R. A. Baker, G. B., & Douglass, A. B. (1991). Depression and memory in major depressive disorder. *Journal of Neuropsychiatry and Clinical Neurosciences, 3(1),* 78-80. DOI: 10.1176/jnp.3.1.78

Bridges, A. J., & Holler, K. A. (2007). How many is enough? Determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology, 13(6),* 528-538. DOI: 10.1080/09297040701233875

Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning and Memory*, *13*, 382–387. DOI: 10.1101/lm.206306

Butler, C. R., Graham, K. S., Hodges, J. R., Kapur, N., Wardlaw, J. M., & Zeman, A. Z. (2007). The syndrome of transient epileptic amnesia. *Ann Neurol, 61,* 587-598.

Butler, C. R., & Zeman, A. Z. (2008). Recent insights into the impairment of memory in epilepsy: Transient epileptic amnesia, accelerated long-term forgetting and remote memory impairment. *Brain, 131,* 2243-2263. DOI: http://dx.doi.org/10.1093/brain/awn127

Butler, C. R., Bhaduri, A., Acosta-Cabronero, J., Nestor, P. J., Kapur, N., Graham, K. S., Hodges, J. R., et al. (2009). Transient epileptic amnesia: Regional brain atrophy and its relationship to memory deficits. *Brain, 132(2),* 357-368.

Cahill, L., & Alkire, M. T. (2003). Epinephrine enhancement of human memory consolidation: Interaction with arousal at encoding. *Neurobiology of Learning and Memory*, *79*, 194–198. DOI: 10.1016/S1074-7427(02)00036-9

Cahill, L., Gorski, L., & Le, K. (2003). Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learning and Memory, 10,* 270-274. DOI: 10.1101/lm.62403

Carlesimo, G. A., Sabbadini, M., Loasses, A., & Caltagirone, C. (1997). Forgetting from long-term memory in severe closed-head injury patients: Effect of retrieval conditions and semantic organisation. *Cortex, 33(1),* 131-142. DOI: 10.1016/S0010-9452(97)80009-3

Carpay, J. A., Aldenkamp, A. P., & van Donselaar, C. A. (2005). Complaints associated with the use of antiepileptic drugs: results from a community-based study. *Seizure, 14,* 198-206.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13,* 181-197.

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion, 7(4),* 802-811. DOI: 10.1037/1528-3542.7.4.802

Corcoran, R., & Thompson, P. (1992). Memory failure in epilepsy: retrospective reports and prospective recordings*. Seizure, 1,* 37-42.

Coccoz, V., Sandoval, A. V., Stehberg, J., & Delorenzi, A. (2013). The temporal dynamics of enhancing a human declarative memory during reconsolidation. *Neuroscience, 246,* 397-408. DOI: 10.1016/j.neuroscience.2013.04.033

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement, 20,* 37–46.

Convit, A., Wolf, O. T., Tarshish, C., & de Leon, M. J. (2003). Reduced glucose tolerance is associated with poor memory performance and hippocampal atrophy among normal elderly. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 100(4),* 2019-2022. DOI: 10.1073/pnas.0336073100

Coste, S., Ryvlin, P., Hermier, M., Ostrowsky, K., Adeleine, P., Froment, J. C., & Mauguière, F. (2002). Temporopolar changes in temporal lobe epilepsy: a quantitative MRI-based study. *Neurology, 59,* 855-61.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for

quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist, 27,* 1173-1195.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12(4),* 482-486. DOI: 1385-4046/98/1204-482

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., & Wilcox, R. R. (2012). Effect of non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical Psychology, 65,* 56-73. DOI: 10.1111/j.2044-8317.2011.02014.x

Darlington, R. B. (1970). Is kurtosis really 'peakedness'? *The American Statistician, 24(2),* 19-22.

Davis, H. P., Small, S. A., Stern, Y., Mayeux, R., Feldstein, S. N., Keller, F. R. (2003). Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex, 39,* 1063-1091.

DeSouza, R. (2004). Motherhood, migration and methodology. *The Qualitative Report, 9(3),* 463-482.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. DOI: 10.1146/annurev.psych.55.090902.142050

Educational Recording Agency. (2011). *A guide to using Broadcasts for Educational Purposes.* London, UK: Author.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(*4*), 277. doi: 10.1027/1016-9040.13.4.277

Eflikides, A., Yiultsi, E., Kangellidou, T., Kounti, F., Dina, F., & Tsolaki, M. (2002). Wechsler Memory Scale, Rivermead Behavioural Memory Test and

Everyday Memory Questionnaire in healthy adults and Alzheimer's patients. *Eur J Psychological Assess, 18,* 63-77.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7(1),* 1–26. DOI:10.1214/aos/1176344552.

Emilien, G., Durlach, C., Atoniadis, E., Van der Linden, M., Maloteaux, J. (2004). *Memory: Neuropsychological, Imaging and Psychopharmacological Perspectives.* Hove, East Sussex: Psychology Press.

Erickson, K. I., Prakash, R. S., Voss, M. W., Chaddock, L., Hu, L., Morris, K. S., White, S. M., Wojcicki, T. R., McAuley, E., & Kramer, A. F. (2009). Aerobic fitness is associated with hippocampal volume in elderly humans. *Hippocampus, 19(10),* 1030–1039. DOI: 10.1002/hipo.20547

Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., Kim, J. S., Heo, S., Alves, H., White, S. M., Wojcicki, T. R., Mailey, E., Vieira, V. J., Martin, S. A., Pence, B. D., Woods, J. A., McAuley, E., & Kramer, A. F. (2011). Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 108(7),* 3017-3022. DOI: 10.1073/pnas.1015950108

Fisher, R. S., Vickrey, B. G., Gibson, P., Hermann, B., Penovich, P., Scherer, A., & Walker, S. (2000). The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. *Epilepsy Res, 41,* 39-51.

Fitzgerald, Z., Mohamed, A., Ricci, M., Thayer, Z., & Miller, L. (2013a). Accelerated long-term forgetting: A newly identified memory impairment in epilepsy. *Journal of Clinical Neuroscience, 20(11);* 1486-1491.

Fitzgerald, Z., Thayer, Z., Mohamed, A., & Miller, L. (2013b). Examining factors related to accelerated long-term forgetting in epilepsy using ambulatory EEG monitoring. *Epilepsia, 54(5),* 819-827.Fossati, P., Harvey, P., Le

Bastard, G., Ergis, A., Jouvent, R., & Allilaire, J. (2004). Verbal memory performance of patients with a first depressive episode and patients with unipolar recurrent depression. *Journal of Psychiatric Research, 38(2),* 137-144. DOI: 10.1016/j.jpsychires.2003.08.002

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R.J. Sbordone, & C.J. Long (Eds.). Ecological validity of neuropsychological testing (pp. 91-112). Boca Raton, FL: St Lucie Press.

Gabrieli, J. D. E., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science, 276(5310),* 264-266.

Gallace, A., & Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychological Bulletin, 135(3),* 380-406. DOI: http://dx.doi.org/10.1037/a0015325

Gilliam, F., Kuzniecky, R., Meador, K., Martin, R., Sawrie, S., Viikinsalo, M., Morawetz, R., & Faught, E. (1999). Patient-oriented outcome assessment after temporal lobectomy for refractory epilepsy. *Neurology, 53,* 687-694.

Giovagnoli, A. R., Casazza, M., & Avanzini, G. (1995). Visual learning on a selective reminding procedure and delayed recall in patients with temporal lobe epilepsy. *Epilepsia, 36,* 704-711.

Gold, J. J., & Trauner, D. A. (2014). Hippocampal volume and memory performance in children with perinatal stroke. *Pediatric Neurology, 50(1),* 18-25. DOI: 10.1016/j.pediatrneurol.2013.08.029

Green, R. E., Melo, B., Christensen, B., Ngo, L. A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of*

Clinical and Experimental Neuropsychology, *30*(2), 163–172. DOI: 10.1080/13803390701300524

Hall, K. E., Isaac, C. L., & Harris, P. (2009). Memory complaints in Epilepsy: An accurate reflection of memory impairment or an indicator of poor adjustment? A review of the literature. *Clinical Psychology Review, 29,* 354-367.

Helmstaedter, C., & Elger, C. E. (2000). Behavioral markers for self- and other-attribution of memory: A study in patients with temporal lobe epilepsy and healthy volunteers. *Epilepsy Research, 41,* 235-243.

Helmstaedter, C., Hauff, M., & Elger, C. E. (1998). Ecological validity of list-learning tests and self-reported memory in healthy individuals and those with temporal lobe epilepsy. *Journal of Clinical & Experimental Neuropsychology, 20(3),* 365-75. DOI: 10.1076/jcen.20.3.365.824

Hendriks, M. P. H., Aldenkamp, A. P., Alpherts, W. C. J., Ellis, J., Vermeulen, J., & Van der Vlugt, H. (2004). Relationships between epilepsy-related factors and memory impairment. *Acta Neurologica Scandinavica, 110(5),* 291-300. DOI: 10.1111/j.1600-0404.2004.00319.x

Herrman, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin, 92(2),* 434-452.

Hertzog, C., & Pearman, A. (2014). *Memory Complaints in Adulthood and Old Age.* In T. J. Perfect, & D. S. Lindsay (Eds.), The SAGE Handbook of Applied Memory (pp. 423-443). London, UK: SAGE Publications Ltd.

Hickox, A., & Sunderland, A. (1992). *Questionnaire and checklist approaches to assessment of everyday memory problems.* In J. R. Crawford, & D. M. Parker (Eds.), A Handbook of Neuropsychological Assessment (pp. 103-114). Mahwah, NJ: Lawrence Erlbaum Associates.

Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in Multiple Sclerosis. *Archives of Clinical Neuropsychology, 15(3),* 185-204. DOI: 10.1016/S0887-6177(99)00004-9

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric Statistical Methods* (3rd ed.). Hoboken: NJ: John Wiley & Sons.

Honea, R. A., Thomas, G. P., Harsha, A., Anderson, H. S., Donnelly, J. E., Brooks, W. M., & Burns, J. M. (2009). Cardiorespiratory fitness and preserved medial temporal lobe volume in Alzheimer's disease. *Alzheimer Disease & Associated Disorders, 23(3),* 188–197.

Howard, C. E., Andrés, P., Broks, P., Noad, R., Sadler, M., Coker, D., & Mazzoni, G. (2010). Memory, metamemory and their discussion in temporal lobe epilepsy. *Neuropsychologia, 48,* 921-932.

Jambaqué, I., Dellatolas, G., Dulac, O., Ponsot, G., & Signoret, J. L. (1993). Verbal and visual memory impairment in children with epilepsy. *Neuropsychologia, 31(12),* 1321-1337. DOI: 10.1016/0028-3932(93)90101-5

Jansari, A. S., Davis, K., McGibbon, T., Firminger, S., & Kapur, N. (2010). When 'long-term memory' no longer means 'forever': Analysis of accelerated long-term forgetting in a patient with temporal lobe epilepsy. *Neuropsychologia, 48(6),* 1707-1715.

Jokeit, H., Daamen, M., Zang, H., Jansky, J., & Ebner, A. (2001). Seizures accelerate forgetting in patients with left-sided temporal lobe epilepsy. *Neurology, 57(1),* 125-126.

Jones, E., Sünram-Lea, S. I., & Wesnes, K. A. (2012). Acute ingestion of different macronutrients differentially enhances aspects of memory and attention in

healthy young adults. *Biological Psychology, 89(2),* 477-486. DOI: 10.1016/j.biopsycho.2011.12.017

Kaplan, R. J., Greenwood, C. E., Winocur, G., & Wolever, T. M. S. (2000). Cognitive performance is associated with glucose regulation in healthy elderly persons and can be enhanced with glucose and dietary carbohydrates. *American Society for Clinical Nutrition, 72(3),* 825-836.

Kaplan, R. J., Greenwood, C. E., Winocur, G., & Wolever, T. M. S. (2001). Dietary protein, carbohydrate, and fat enhance memory performance in the healthy elderly. *American Society for Clinical Nutrition, 74(5),* 687-693.

Keppel, G., & Wickens, T. D. (2004). Design and analysis: A researcher's handbook (4th ed.). New Jersey, USA: Pearson.

Kim, H. (2012). Statistical notes for clinical researchers: assessing normal distribution (1). *Restorative Dentistry & Endodontics, 37(4),* 245-248. DOI: 10.5395/rde.201.37.4.245

Kim, H. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38(1),* 52-54. DOI: 10.5395/rde.2013.38.1.52

Kim, H. (2014a). Statistical notes for clinical researchers: Nonparametric statistical methods: 1. Nonparametric methods for comparing two groups. *Restorative Dentistry & Endodontics, 39(3),* 235-239. DOI: 10.5395/rde.2014.39.3.235

Kim, H. (2014b). Statistical notes for clinical researchers: Nonparametric statistical methods: Nonparametric methods for comparing three or more groups and repeated measures. *Restorative Dentistry & Endodontics, 39(4),* 329-332. DOI: 10.5395/rde.2014.39.4.329

Kirk-Sanchez, N. J., & McGough, E. L. (2014). Physical exercise and cognitive performance in the elderly: current perspectives. *Clinical Interventions in Aging, 9,* 51-62. DOI: 10.2147/CIA.S39506

Kizilbash, A. H., Vanderploeg, R. D., & Curtiss, G. (2002). The effects of depression and anxiety on memory performance. *Archives of Clinical Neuropsychology, 17(1),* 57-67. DOI: 10.1016/S0887-6177(00)00101-3

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics and Data Analysis, 51,* 5731-5742.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Langdridge, D. (2004). *Introduction to research methods and data analysis in psychology.* Harlow, Essex: Pearson Education Ltd.

Levin, H. S., High Jr., W. M., & Eisenberg, H. M. (1988). Learning and forgetting during posttraumatic amnesia in head injured patients. *Journal of Neurology, Neurosurgery & Psychiatry, 51,* 14-20. DOI: 10.1136/jnnp.51.1.14

Lewis, P., & Kopelman, M. D. (1998). Forgetting rates in neuropsychiatric disorders. *Journal of Neurology, Neurosurgery & Psychiatry, 65,* 890-898. DOI: 10.1136/jnnp.65.6.890

Lucchelli, F., & Spinnler, H. (1998). Ephemeral new traces and evaporated remote engrams: A form of neocortical temporal lobe amnesia? *Neurocase, 4,* 447-459.

Lupien, S. J., Gaudreau, S., Tchiteya, B. M., Maheu, F., Sharma, S., Nair, N. P., Hauger, R. L., McEwen, B. S., & Meaney, M. J. (1997). Stress-induced declarative memory impairment in healthy elderly subjects: relationship to

cortisol reactivity. *The Journal of Clinical Endocrinology and Metabolism, 82(7),* 2070-2075. DOI: http://dx.doi.org/10.1210/jcem.82.7.4075

Mameniskiene, R., Jatuzis, D., Kaubrys, G., & Budrys, V. (2006). The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. *Epilepsy & Behavior, 8(1),* 278-288. DOI: 10.1016/j.yebeh.2005.11.003

Manning, C. A., Hall, J. L., & Gold, P. E. (1990). Glucose effects on memory and other neuropsychological tests in elderly humans. *Psychol Sci, 1,* 307-311.

Manning, C. A., Parsons, M. W., Cotter, E. M., & Gold, P. E. (1997). Glucose effects on declarative and nondeclarative memory in healthy elderly and young adults. *Psychobiology, 25,* 103-108.

Martin, R. C., Loring, D. W., Meador, K. J., Lee, G. P., Thrash, N., & Arena, J. G. (1991). Impaired long-term retention despite normal verbal learning in patients with temporal lobe dysfunction. *Neuropsychology, 5(1),* 3.

Mary, A., Schreiner, S., & Peigneux, P. (2013). Accelerated long-term forgetting in aging and intra-sleep awakenings. *Frontiers in Psychology, 4,* 1-11. DOI: 10.3389/fpsyg.2013.00750

McAuley, J. W., Elliott, J. O., Patankar, S., Hart, S., Long, L., Moore, J. L., & Shneker, B. F. (2010). Comparing patients' and practitioners' views on Epilepsy concernes: A call to address memory concerns. *Epilepsy & Behavior, 19,* 580-583.

McDaniel, M. A., & Masson, M. E. J. (1978). *Organizational factors in delayed recall and recognition. Program on Cognitive Factors in Human Learning and Memory Report No. 78.* Colorado, USA: Institute for the Study of Intellectual Behavior, University of Colorado.

McGeoch, J. A. (1942). *The psychology of human learning.* New York, USA: Longman.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(*3*), 276–282. http://dx.doi.org/10.11613/BM.2012.031

Meikle, A., Riby, L. M., & Stollery, B. (2004). The impact of glucose ingestion and gluco-regulatory control on cognitive performance: a comparison of younger and middle aged adults. *Human Psychopharmacology: Clinical and Experimental, 19(8),* 523-535.

Messier, C., Desrochers, A., & Gagnon, M. (1999). Effect of glucose, glucose regulation, and word imagery value on human memory. *Behavioral Neuroscience, 113(3),* 431-438. DOI: http://dx.doi.org/10.1037/0735-7044.113.3.431

Millen, D. (1997). Some methodological and epistemological issues raised by doing feminist research on non-feminist women. *Sociological Research On-line, 2(3).*

Motamedi, G. K., & Meador, K. J. (2004). Antiepileptic drugs and memory. *Epilepsy & Behavior, 5(4),* 435-439.

Muhlert, N., Grünewald, R. A., Hunkin, N. M., Reuber, M., Howell, S., Reynders, H., & Isaac, C. L. (2011). Accelerated long-term forgetting in Temporal Lobe but not Idiopathic Generalised Epilepsy. *Neuropsychologia*, *49(9)*, 2417–2426.

Muhlert, N., Milton, F., Butler, C. R., Kapur, N., & Zeman, A. Z. (2010). Accelerated forgetting of real-life events in Transient Epileptic Amnesia. *Neuropsychologia*, *48*(11), 3235–3244.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7(2),* 217-227.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000), Multiple trace theory of human memory: Computational, neuroimaging, and

neuropsychological results. *Hippocampus, 10(4),* 352–368. DOI: 10.1002/1098-1063

Nelson, H. E. (1982). *The National Adult Reading Test (NART): test manual*. NFER-Nelson.

Pallant, J. (2010). *SPSS survival manual* (4th ed.). Berkshire, UK: Open University Press.

Piazzini, A., Canevini, M. P., Maggiori, G., & Canger, R. (2001). The perception of memory failures in patients with epilepsy. *European Journal of Neurology, 8(6),* 613-20. DOI: 10.1046/j.1468-1331.2001.00287.x

Podewils, L. J., Guallar, E., Kuller, L. H., Fried, L. P., Lopez, O. L., Carlson, M., & Lyketsos, C. G. (2005). Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *Am J Epidemiol., 161(7),* 639-651.

Roozendaal, B. (2002). Stress and memory: Opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory*, *78*(3), 578–595. DOI: 10.1006/nlme.2002.4080

Rosner, S. R. (1971). The effects of rehearsal and chunking instructions on children's multitrial free recall. *Journal of Experimental Child Psychology, 11(1);* 93-105.

Ryan, L., & Golden, A. (2006). 'Tick the box please': a reflexive approach to doing quantitative social research. *Sociology, 40(6),* 1191-1200.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R.J. Sbordone, & C.J. Long (Eds.). Ecological validity of neuropsychological testing (pp. 15-41). Boca Raton, FL: St Lucie Press.

Seier, E. (2002). Comparison of tests of univariate normality. *Interstat. Statistical Journal, 1,* 1-17.

Small, G. W. (2002). What we need to know about age related memory loss. *British Medical Journal, 324(7352)*, 1502-1505.

Sparrow, B., Liu, J., Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 5(333);* 776-778. DOI: 10.1126/science.1207745

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology, 21(4),* 327-337. DOI: 10.1016/j.acn.2006.04.004

Squire, L.R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology, 5,* 169-l77.

Squire, L. R., & Bayley, P. J. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology, 17,* 185-196. DOI: 10.1016/j.conb.2007.02.006

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience, 27(1),* 279–306. DOI: 10.1146/annurev.neuro.27.070203.144130

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253,* 1380-1386.

Stockburger, D. W. (2013). *Introductory statistics: Concepts, models, and applications* (3rd ed.). Missouri, USA: Missouri State University.

Stroth, S., Hille, K., Spitzer, M., & Reinhardt, R. (2009). Aerobic endurance exercise benefits memory and affect in young adults. *Neuropsychological*

Rehabilitation: An International Journal, 19(2), 223-243. DOI: 10.1080/09602010802091183

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22,* 341-357.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1984). Assessing every-day memory after severe head injury. In J. E. Harris & P. E. Morris (Eds.), Everyday memory, actions, and absent-mindedness (pp. 193-212). London, UK: Academic Press.

Sunderland, A., Harris, J. E., & Gleave, J. (1984). Memory failures in everyday life following severe head injury. *J Clin Neuropsychol, 6,* 125-141.

Tabachnik, B., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). New Jersey, USA: Pearson.

Téllez-Zenteno, J. F., Dhar, R., Hernandez-Ronquillo, L., & Wiebe, S. (2007). Long-term outcomes in epilepsy surgery: antiepileptic drugs, mortality, cognitive and psychosocial aspects. *Brain, 130(2),* 334-345. DOI: http://dx.doi.org/10.1093/brain/awl316

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience, 100,* 147-154.

Tollenaar, M. S., Elzinga, B. M., Spinhoven, P., & Everaerd, W. A. M. (2008). The effects of cortisol increase on long-term memory retrieval during and after acute psychosocial stress. *Acta Psychologica*, *127*, 542–552. DOI: 10.1016/j.actpsy.2007.10.007

Trammell, J. P., & Clore, G. L. (2013). Does stress enhance or impair memory consolidation? *Cognition and Emotion, 28(2),* 361-374. DOI: 10.1080/02699931.2013.822346

Troyer, A. K., & Rich, J. B. (2002). Psychometric properties of a new metamemory questionnaire for older adults. *The Journals of Gerontology: Series B, 57(1),* 19-27. DOI: 10.1093/geronb/57.1.P19

Tu, S., Miller, L., Piguet, O., & Hornberger, M. (2014). Accelerated forgetting of contextual details due to focal medio-dorsal thalamic lesion. *Frontiers in Behavioral Neuroscience, 8,* 1-9. DOI: 10.3389/fnbeh.2014.00320

Tu, S., Mioshi, E., Savage, S., Hodges, J. R., & Hornberger, M. (2013). Dissociation of explicit and implicit long-term memory consolidation in semantic dementia: A case study. *Neurocase: The Neural Basis of Cognition, 19(4),* 401-407. DOI: 10.1080/13554794.2012.690424

Uijl, S. G., Uiterwaal, A. P., Aldenkamp, A. P., Carpay, J. A., Doelman, J. C., Keizer, K., VEcht, C. J., de Krom, M. C., & van Donselaar, C. A. (2006). A cross-sectional study of subjective complaints in patients with epilepsy who seem to be well-controlled with anti-epileptic drugs. *Seizure, 15,* 242-248.

Von Hippel, P. T. (2005). Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education, 13(2).*

Walker, S., Read, S., & Priest, H. (2013). Use of reflexivity in a mixed-methods study. *Nurse Researcher, 20(3),* 38-43.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Pittsburgh, PA: Springer

Weingartner, H., & Parker, E. S. (Eds.). (2014). *Memory Consolidation: Psychobiology of Cognition.* Hove, East Sussex: Psychology Press.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2009). *Wechsler Memory Scale–Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Wickelgren, W. A. (1979). Chunking and consolidation: a theoretical synthesis of semantic networks, configuring, S-R versus cognitive learning, normal forgetting, the Amnesic Syndrome, and the hippocampal arousal system. *Psychological review, 86,* 44-60.

Wilkinson, H., Holdstock, J. S., Baker, G., Herbert, A., Clague, F., & Downes, J. J. (2012). Long-term accelerated forgetting of verbal and non-verbal information in temporal lobe epilepsy. *Cortex, 48,* 317-332.

Winter, B., Breitenstein, C., Mooren, F.C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A., & Knecht, S. (2007). High impact running improves learning. *Neurobiology of Learning and Memory,* 87(4), 597–609.

Extended Appendices

Appendix H. Copy of REC approval email

RE: Thesis ethics
Aidan Hart
Sent: 07 March 2014 09:53
To: Emma Cameron (12353908)
Cc: Patrick Bourke

Hi Emma,

All this looks fine.

I am happy to confirm that the concerns and queries of the first and second reviewer have now been addressed and I am happy to inform you that your study has ethical approval to proceed. I hope it goes well

With best wishes

Aidan

Dr. Aidan Hart CPsychol (Clinical/Forensic)
HCPC Registered Clinical and Forensic Psychologist
Academic Tutor
Trent Doctoral Programme in Clinical Psychology
University of Lincoln
Lincoln LN6 7TS

01522 886029

Appendix I. Participant Information Sheet

UNIVERSITY OF LINCOLN

PARTICIPANT INFORMATION SHEET

Title of study: study of the psychometric properties of the ALFIE, a novel measure of accelerated long-term forgetting in temporal lobe epilepsy

Name of researcher: Emma Cameron

I would like to invite you to take part in my research study. Before you decide I would like you to understand why the research is being done and what it would involve for you. I will go through this participant information sheet with you and answer any questions you have. Talk to others about the study if you wish. Ask me if there is anything that is not clear.

What is the purpose of the study?

The purpose of this study is to continue the development of a memory test for those with temporal lobe epilepsy. The type of memory being looked at is known as accelerated long-term forgetting. At present, memory tests are not sufficiently designed to allow problems with accelerated long-term forgetting to be shown. Therefore, when people with temporal lobe epilepsy complain of memory difficulties, their problems do not often show in tests. By continuing the development of this test (the ALFIE) we hope to combat this. Furthermore, we hope that our test will be more cost-effective, and more like real-life, than alternative measures.

Why have I been invited?

You are being invited to take part because you are part of a healthy population sample. We are inviting an estimated 50 participants like you to take part. By understanding how healthy participants perform on the memory test we can create statistical norms for the test.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this participant information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This means you could withdraw at any point during the study or up to two weeks after completing the study. This would not affect your legal rights.

What will happen to me if I take part?

If you decide to take part the study takes place in two stages:

Stage 1: Complete some memory measures and short questionnaires. This will include remembering things immediately and after a short delay. This could take around an hour. This will take place on the University of Lincoln campus/your local group's meeting place/your home depending on how you were contacted by the researcher.

Stage 2: The researcher will contact you by phone two weeks later and ask you to do some other tests and ask you some more questions. This could take around 30-40 minutes.

Between stages 1 and 2 you will not have to do anything. After stage 2 your participation will be complete.

Expenses and payments

Participants will not be paid to participate in the study and unfortunately we are unable to reimburse any travel expenses; however, as a thank you for taking part you have the option to enter into a prize-draw for a £50.00 voucher for a high street shop. If you wish to do so, your email address/postal address will need to be taken for the purposes of contacting you if you win the prize-draw.

What are the possible disadvantages and risks of taking part?

There are no anticipated risks associated with this study; however participants should be aware that they may not perform as well as they expected; if you have any concerns you can speak with the researcher.

What are the possible benefits of taking part?

We cannot promise the study will help you but the information we get from this study may help those with temporal lobe epilepsy, whose memory difficulties do not currently show on tests. In the future, this measure may be expanded to other clinical populations who also experience memory problems, for example traumatic brain injuries.

As a result, it may help referrals for those who need memory rehabilitation, or help inform interventions in psychological therapies.

What happens when the research study stops?

After Stage 2 of the research your involvement will be complete. If you wish to withdraw from the study, you may do so at any point during your participation and up to two weeks after you have completed Stage 2 of the research (see section below '*What will happen if I don't want to carry on with the study?*' for more information on withdrawing).

The data provided by participants will be compared at different time points and by certain demographics to try to ascertain how the healthy population would perform on this test. This will be written up as part of the Trent DClinPsy doctoral thesis and may be disseminated. Participants will not be identified in any publications.

If you wish to receive a summary of the research findings once the study has been completed, you may do so. The researcher will ask you if you would like to leave an email address for this purpose.

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. If you remain unhappy and wish to complain formally, you can do this by contacting the University of Lincoln. Details can be obtained from the university. Or, if you feel that the research has ethical issues that need to be reviewed, you may contact the University of Lincoln's Research Ethics Committee if you wish. All contact details are provided at the end of this information sheet.

Will my taking part in the study be kept confidential?

We will follow ethical and legal practice and all information about you will be handled in confidence.

If you join the study, some parts of the data collected for the study will be looked at by authorised persons from the University of Lincoln who are organising the research. They may also be looked at by authorised people to check that the study is being carried out correctly. All will have a duty of confidentiality to you as a research participant.

All information which is collected about you during the course of the research will be kept **strictly confidential**, stored in a secure and locked office, and on a password protected database. Any data transferred from one office to another (e.g. for analysis) will be done so securely on an encrypted USB stick.

You will be assigned a code number for use on any study documents and the electronic database. The code number will be constructed from your initials, an allocated study number and the first four numbers from their DOB, e.g. if your details were;
Name: Emma Cameron, DOB: 15.03.1955 and you were assigned a study number of 001 then your code number would look like this: 001_EC_1503. This means that data from tests is non-identifiable by the researchers.

Your personal contact data (e.g. telephone number, email address) will be kept until you have been contacted for stage two of the study. After this point it will be destroyed. The only exception to this is if you have requested a summary of the research findings and/or wish to enter into the prize-draw, in which case your name and email address/postal address will need to be kept to contact you about these. After contact, this information will be destroyed. Once the research is written up your contact name linking you to your participant code will also be destroyed. All other data (research data) will be kept securely for 7 years. After this time your data will be disposed of securely. During this time all precautions will be taken by all those involved to maintain your confidentiality; only members of the research team will have access to your personal data. These procedures for handling, processing, storage and destruction of data meets the requirements of the Data Protection Act 1998.

What will happen if I don't want to carry on with the study?

Your participation is voluntary and you are free to withdraw at any time, without giving any reason, and without your legal rights being affected.
You may withdraw at any point during your participation (i.e. at Stage 1, Stage 2, or in-between these two stages) and up until two weeks after you have finished

Stage 2 of the research, by contacting the researcher to discuss this. The two week withdrawal date will be given to you on the telephone after Stage 2 has been completed.

If you decide to withdraw during the study or during the two week withdrawal period after the study, all of your data will be removed and destroyed. You will need to provide the researcher with your unique participant code number that you create at Stage 1 of the research in order that the researcher can locate your data, as all data is automatically anonymised.

Involvement of the General Practitioner/Family doctor (GP)

Your GP does not need to be notified of your participation in this study.

Who is organising and funding the research?

This research is being organised and funded by the University of Lincoln as part of the Trent DClinPsy programme.

Who has reviewed the study?

All research is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given approval by the University of Lincoln's Research Ethics Committee.

Further information and contact details

Name	Capacity	Contact Details
Emma Cameron, Trainee Clinical Psychologist	Principal Investigator and Statistician	Address: Faculty of HLSS University of Lincoln 1st Floor, Bridge House Brayford Pool Lincoln LN6 7TS
Nima Moghaddam, Research Tutor	Primary Chief Investigator, Research Supervisor and Study Statistician	Address: Faculty of HLSS University of Lincoln 1st Floor, Bridge House Brayford Pool Lincoln LN6 7TS
Roshan Das Nair, Research Tutor	Secondary Chief Investigator, Research Supervisor	Address: Institute of Work Health and Organisations International House Jubilee Campus University of Nottingham Wollaton Road Nottingham NG8 1BB

Study Coordinating Centre		Address: Faculty of HLSS University of Lincoln 1st Floor, Bridge House Brayford Pool Lincoln LN6 7TS
Patrick Bourke	Chair of the Research Ethics Committee	Address: Faculty of HLSS University of Lincoln 1st Floor, Bridge House Brayford Pool Lincoln LN6 7TS

Appendix J. Consent form

UNIVERSITY OF LINCOLN

CONSENT FORM

Title of Study: A study of the psychometric properties of the ALFIE, a novel measure of accelerated long-term forgetting in temporal lobe epilepsy

REC ref: _____

Name of Researcher: _____

Name of Participant: _____ **Please initial boxes**

1. I confirm that I have read and understand the participant information sheet version number ……dated..... for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time up until two weeks after my completion of Stage 2 of the research, without giving any reason, and without my medical care or legal rights being affected. I understand that if I wish to withdraw after this period of time then the information collected cannot be erased and this information can still be used. ☐

3. I understand that the data collected in the study may be looked at by authorised individuals from the University of Lincoln, the research group and regulatory authorities where it is relevant to my taking part in this study. I give permission for these individuals to have access to these records and to collect, store, analyse and publish information obtained from my participation in this study. I understand that my personal details will be kept confidential. ☐

4. I agree to take part in the above study. ☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from Principal Investigator)	Date	Signature
_____	_____	_____
Name of Principal Investigator	Date	Signature

2 copies: 1 for participant, 1 for the project notes

Appendix K. Contact sheet

CONTACT SHEET
Participant code number: _____

Section 1:
This study takes part in two stages. To be able to carry out stage 2, the researcher will need to be able to contact you by telephone. This can be a mobile telephone or a landline. If you could therefore fill out the following details:

Name (in block capitals):

_____

Telephone Number: ☐☐☐☐☐☐☐☐☐☐☐

Please note, once your involvement in the study is complete, the above information will be destroyed.

- -

Section 2:
You only need to fill in this section of the sheet if you wish to:
(a) Receive a summary of the results of the study once completed, and/or
(b) Wish to be entered into the prize-draw.

Please tick below to inform the researcher why you are leaving your contact details. You may tick one or both options.

I wish to receive a summary of the results of the study once completed ☐

I wish to be entered into the prize-draw ☐

Name (in block capitals):

_____

Email address (please write clearly):

_____

Please note:
If you have filled out Section 2 of this form then your name and email address will be transferred to a separate confidential document. This means it will be kept separate to any other data held by the researcher and will not be able to be linked in any way to any data gathered. This document will be encrypted and password protected. Once you have been contacted with the results of the study and/or the prize-draw has been drawn, this information will be destroyed.

Contact regarding the results and the prize-draw can only be made via email address due to the funding limitations of the study.

Entry into the prize-draw is made on completion of Stage 2 of the research. If you are not successful in the prize-draw, you will not hear from the researcher. Only the successful person will be contacted.

Appendix L. Homogeneity of variance tables

Table 23. Equality of covariance matrices and of error variances for BMIPB recall tasks

Objective memory measure	Test version	Between-groups variable	Levene's F	Ratio rule
BMIPB story recall	V1 & V2	Gender	T1-T3: F(1,48)=.01-1.08, p≥.30	
		Age	T1: F(8,41)=2.31, p=.04 T2-T3: F(8,41)=.58-2.04, p≥.07	T1: 1:4
		Years of education	T1: F(4,45)=5.02, p=.01 T2: F(4,45)=4.42, p=.01 T3: F(4,45)=2.90, p=.03	T1-T3: 1:2
BMIPB list recall	V1	Gender and years of education	T1-T3: F(1,21)=.16-2.37, p≥.10	
		Age	T1&T3: F(6,16)=1.05-1.29, p≥.32 T2: F(6,16)=4.60, p=.01	T2: 1:4
	V2	Gender	T1-T3: F(1,19)=.05-.39, p≥.54	
		Age	T1: F(5,15)=4.69, p=.01 T2-T3: F(5,15)=1.34-1.83, p≥.17	T1: 1:4
		Years of education	T1-T2: F(3,17)=1.79-2.96, p≥.06 T3: F(3,17)=3.77, p=.03	T3: 1:2

BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay.

Table 24. Equality of covariance matrices and of error variances for BMIPB recognition tasks

Objective memory measure	Test version	Between-groups variable	Levene's F	Ratio rule
BMIPB List: list recognition	V1	Gender and years of education	T2-T3: F(1,21)=.03-1.14, p≥.36	
		Age	T2: F(6,16)=2.47, p=.07	
			T3: F(6,16)=12.31, p=.01	T3: 1:5
	V2	Gender	T2-T3: F(1-19)=.04-1.47, p≥.24	
		Age	T2: F(5,15)=3.09, p=.04	T2: 1:2
			T3: F(5,15)=1.39, p=.28	
		Years of education	T2: F(3,17)=9.85, p=.01	T2: 1:8
			T3: F(3,17)=1.16, p=.36	
BMIPB List: word recognition	V1	Gender, age and years of education	T2-T3: F(1,21)=.72-2.21, p≥.10	
	V2	Gender, age and years of education	T2-T3: F(1,19)=.16-2.47, p≥.08	

BMIPB = Brain Injury Rehabilitation Trust Memory and Information Processing Battery (Coughlan, Oddy, & Crawford, 2007); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay.

Table 25. Equality of covariance matrices and of error variances for ALFIE recall tasks

Objective memory measure	Test version	Between-groups variable	Levene's F	Ratio rule
ALFIE story	V1	Gender and years of education	T1-T3: F(1,21)=.01-1.98, p≥.15	
		Age	T1-T2: F(6,16)=2.20-2.65, p≥.06	
			T3: F(6,16)=4.75, p=.01	T3: 1:2
	V2	Gender	T1-T3: F(1,19)=.26-1.33, p≥.26	
		Age	T1: F(5,15)=3.68, p=.02	T1: 1:10
			T2-T3: F(5,15)=.37-2.69, p≥.06	
		Years of education[a]	T1-T2: F(3,17)=.20-1.05, p≥.40	
			T3: F(3,17)=9.77, p=.01	T3: 1:49

ALFIE = Accelerated Long-term Forgetting in Epilepsy test (Corbett, 2012); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay; [a]Indicates that Levene's F-statistic was significant at the .01 level and the ratio rule was violated.

Table 26. Equality of covariance matrices and of error variances for ALFIE recognition tasks

Objective memory measure	Test version	Between-groups variable	Levene's F	Ratio rule
ALFIE narrative recognition	V1	Gender and years of education	T2-T3: F(1,21)=.01-2.59, p≥.08	
		Age	T2: F(6,16)=9.05, p=.01	T2: 1:4
			T3: F(6,16)=2.20, p=.10	
	V2	Gender, age and years of education	T2-T3: F(1,19)=.02-2.80, p≥.06	
ALFIE visual recognition	V1	Gender and years of education	T2-T3: F(1,21)=.11-2.60, p≥.08	
		Age	T2: F(6,16)=2.04, p=.12	
			T3: F(6,16)=7.45, p=.01	T3: 1:4
	V2	Gender, age and years of education	T2-T3: F(1.19)=.15-2.62, p≥.09	

ALFIE = Accelerated Long-term Forgetting in Epilepsy test (Corbett, 2012); V1 = Test Version 1; V2 = Test Version 2; T1 = Immediate recall; T2 = 40-minute delay; T3 = 2-week delay.

Appendix M. Pilot 0-1-2 ALFIE score sheet

For ALFIE Story V2:

	Information unit (bold italicised) and associated response guidelines
1	**News from *Sarah*** 2 = Sarah 1 = Whole name replaced by ‘a woman’/ ‘a lady’ 0 = Whole name replaced by ‘a person’ …… [Wrong first name stated]
2	***Outen*** 2 = Outen [pronounced OO-TEN] 1 = Outen [pronounced OUT-TEN] …… Alton …… [rhymes with OO-TEN, e.g. Houton, Luton etc.] 0 = [Wrong surname stated]
3	**The *26 year old*** 2 = 26 1 = In her mid-twenties 0 = [Other age stated …… No age stated]
4	***Has completed the* tricky *first leg of her* record breaking human powered** 2 = Completed/ finished/ managed/ been able to complete/ achieved her first leg/ part/ section/ bit 1 = Completed/ finished/ managed/ been able to complete/ achieved …… Did a leg/ part/ section/ bit/ first half 0 = Started/ was on
5	***Loop of the globe*** 2 = Loop of/ circle the globe …… Round the world trip/ expedition/ journey/ travels …… Circumnavigating the globe 1 = Trip/ expedition/ journey …… Travelling across the world 0 = [no examples]
6	***Two years ago Sarah was awarded an MBE*** 2 = Two years ago she got an MBE 1 = Two years ago …… Two years ago she did a loop of the globe …… She got an MBE 0 = She got an award …… [Other number of years stated]
7	**After becoming the *first woman to row solo across the Indian Ocean*** 2 = First woman/ female to row/ sail / kayak by herself/ solo/ alone/ single-handedly over the Indian Ocean 1 = First person/ only woman to row [solo] across the Indian Ocean …… First woman to row solo over the Atlantic/ [Other name of Ocean stated except Indian Ocean] …… She rowed over/has done the Indian Ocean 0 = She rowed over the ocean/ Atlantic/ [Other name of Ocean stated except Indian Ocean]
8	***Today she arrived in Tokyo after an epic expedition* across land and sea** 2 = Today she arrived/ reached Tokyo after an epic/ amazing/ incredible expedition/ trip/ journey/ travels 1 = She arrived/ reached Tokyo …… She had an epic/ amazing/ incredible expedition 0 = She arrived/ reached/ went to Japan

9	***Sarah Teal* reports** 2 = [The reporter was] Sarah Teal 1 = Sarah …… Teal [First name or surname only] 0 = There was a female reporter
10	**Since *setting off from London*** 2 = Set off/ started/ commenced/ Left from London 1 = Started/ commenced from England/ the UK 0 = [Other place name that started from]
11	***On April the 1st*** 2 = [the date was] April 1st …… April Fool's Day 1 = April …… [1st of another month] …… 1st of the month 0 = [Other date stated]
12	**Sarah Outen has *covered a lot of land and sea*** 2 = She has covered/ travelled/ crossed/ gone over a lot of land and sea/ oceans/ water 1 = Has covered a lot of land …… Has covered a lot of sea 0 = [no examples]
13	***On her bike and in a kayak*** 2 = On her bike/ bicycle and in a kayak/ canoe 1 = On her bike/ bicycle …… In a kayak/ canoe …… On her bike and by rowing …… By cycling and kayaking 0 = By foot …… On a boat …… [Other method of transport]
14	***Seven months on* she's reached Tokyo. It's the end of the first leg of her London to London record breaking attempt to loop the globe.** 2 = Seven months on/ later 1 = Many months on/ later 0 = [Other length of time stated]
15	**It's been *full of adventures and challenges* OR *and highs and lows* and** 2 = It's been full of adventures and challenges/ difficulties …… It's been full of highs and lows/ ups and downs 1 = It's been full of adventures …… It's been tough/ full of challenges …… It's been fun and hard …… There have been highs/ lows 0 = [no examples]
16	**I suppose there were times *when I wasn't quite sure how I was going to make it this far*…um…** 2 = I wasn't quite sure how I was going to make it this far/ as far as this …… I didn't think/ know how I would make it this far 1 = I wasn't going to make it this far …… I wasn't quite sure how I was going to do this/ it …… She didn't always believe she could do it …… Didn't think she'd be able to do all that 0 = I wasn't going to come this far/ go this far
17	**Whether that's boshing through the *heat of the Gobi desert*** 2 = Heat of the Gobi [desert] …… It was [very] hot in the Gobi [desert] 1 = Heat of the desert …… Was in the [Gobi] desert 0 = The heat/ humidity
18	**Or *in thick mud in Russia,*** 2 = thick mud/ lots of mud in Russia 1 = thick mud/ lots of mud/ muddy roads …… conditions in Russia …… went through/ across Russia 0 = [no examples]

19	***Bits going wrong with bits of equipment and so on*…it's been brilliant** 2 = Bits going/ went wrong/ had trouble with some of her equipment/ the equipment 1 = Things went wrong …… Some of her equipment broke/ needed repairs ……Various problems with her bike 0 = Her equipment failed her …… Lots of technical failures
20	**Sarah from *Rutland in Oakham*** 2 = She is from Rutland in Oakham 1 **=** Rutland …… Oakham 0 = [Other place name stated]
21	**Has enjoyed the *highs of the beautiful landscape*** 2 = The beautiful/ breath-taking/ stunning landscape/ scenery/ surroundings 1 = Nice/ pleasant landscape 0 = [no examples]
22	**And *the local wildlife*** 2 = local wildlife …… Local [wild] animals 1 = wildlife/ [wild] animals 0 = [no examples]
23	**I'm about fifty metres away from a *brown bear* [inaudible 'breathe in']** 2 = Brown bear 1 = [Grizzly/ Wild] bear 0 = [Other animal stated]
24	**And endured the *lows of the dangerous roads*** 2 = Dangerous/ perilous/ hazardous/ treacherous roads 1 = Roads weren't very good …… Roads were unsafe/ poor …… Dangerous traffic 0 = [no examples]
25	***And treacherous seas*** 2 = Treacherous/ dangerous/ perilous/ hazardous seas 1 **=** Choppy seas …… The seas were not calm 0 = [no examples]
26	**Since leaving London Sarah has travelled *11,000 miles*,** 2 = 11,000 miles 1 = [Figure between 10,000 and 15,000 miles stated] 0 = Many miles [Other number of miles stated]
27	***Through 12 countries*** 2 = 12 countries 1 = [Figure between 10 and 15 countries stated] 0 = Many countries …… 12 continents
28	***And kayaked 300 nautical miles to reach Japan.*** 2 = Kayaked/ rowed/ canoed/ sailed 300 [nautical] miles to reach Japan/ Tokyo 1 = Kayaked 300 miles …… 300 nautical miles 0 = Kayaked many miles

Total score available = 56

Poster

[PAGE INTENTIONALLY LEFT BLANK]

Small Scale Research and Impact

Summary of Service-Related Research and associated Impact (SSRI)

Trainee(s)	Supervisor(s)	Placement	Cohort	Date Completed
Emma Cameron	Kerry Beckley	FPB	1213	Aug 2014

Research background and context

Jorm (2000) stated that there is poor public understanding of mental health and mental disorders. This is reflected within the personality disorder population with research into staff perceptions suggesting that clients with a personality disorder diagnosis are perceived as more dangerous than other clients and with staff holding less optimism regarding outcomes (Markham, 2003) as clients are deemed more difficult to manage (Newton-Howes, Weaver, & Tyrer, 2008). This research holds true across a range of professionals (e.g., see Markham, 2003; Newton-Howes et al., 2008).

As a national response to this, the Knowledge and Understanding Framework (KUF) programme was commissioned by the Ministry of Justice in 2007. This is a national programme that aims to inform clinicians, other professionals and service users about personality disorder, challenging misperceptions and providing an understanding of the development of the disorder, along with techniques to help treat. The aim of the programme is to "support people to work more effectively with personality disorder" and improve the service user experience (KUF, 2013). There are three levels to the programme, with increasing knowledge and complexity.

Similar to other forensic services across the UK, the Community Forensic Psychology Service (CFS) works with a large number of clients who present with a personality disorder or personality disorder symptomology. It was considered important by the local KUF programme and the CFS to effectively evaluate the Awareness Level 1 package, the most highly attended level of KUF training, to which some of the clients from the CFS and local professionals are referred.

Research aims

The clinical psychologist within the CFS was keen to formally evaluate the effectiveness of the KUF training in order to provide feedback to commissioners related to funding. The programme could then hopefully be amended based on this feedback (and the collation of feedback from other KUF programmes as it is a national framework) regarding any change in individuals perceptions and attitudes, and any subsequent changes staff would be making to their working practise.

What the research discovered

The evaluation forms for the KUF training programme were analysed utilising a predominantly qualitative content analysis method. The data was analysed in several stages as suggested by Graneheim & Lundman (2004). Overall, there appeared to be an even mix of both positive and negative comments regarding the programme, suggesting that professionals appreciated the training package, but felt that there was scope for improvement.

Highlights of the training included the variety of people present; professionals were helpful to foster discussion, reflections and sharing of knowledge, and the presence of a service user facilitator helped to challenge staff perceptions and provide a realistic understanding of personality disorders. The presentation of materials and the multi-modal aspect of the materials was well received and useful to keep staff engaged and interested. However, the quality of the videos was deemed very poor, which did create a barrier to learning, as did technological difficulties related to the online modules, which prevented some online programme features being used. Repetitiveness was also discussed as a negative experience throughout the entire training, with comments that both materials and group work were of a repetitive nature. Some professionals also commented that material was too simplistic and they wished for more in-depth information (this particularly related to the topics of Personality Disorders and Schema Therapy). Furthermore, there was feedback that the exploration of topics during face-to-face sections of the training package could have been managed better to prevent lengthy discussions, negative comments and the breeding of discontent. Finally, there was also a theme that professionals were dissatisfied with the attitude of the Trust, firstly towards training packages, with statements that they were not allowed time to complete the training so had to manage

modules on top of their workload, or complete it during personal time and secondly, with regards to the current level of support and supervision they received.

A further limitation, was the actual evaluation form itself; there was a paucity of questions focused specifically on the face-to-face training days with an over-emphasis on other components. Mixed reviews as to a preference for online or face-to-face training may have been a result of there not being many questions dedicated solely to the later. Furthermore, questions did not specifically distinguish between different training days (as there were a total of three days training) and so it is unknown whether one particular training day/topic was better received than others. The validity of the form was also questionable, as the information was gathered after all of the training was complete, not as each section of training was finished, which led to a potential for retrospective bias by participants. In addition, the evaluation forms were not anonymised; asking for the name, role and organisation of the individual filling it out. This means that response biases are more likely to have occurred, although a mix of positive and negative results, particularly for the qualitative sections of the evaluation form, do suggest that this was not an issue and that it was a valid data set. Furthermore, some of the questions did not lend themselves well to the construct they were supposed to be measuring and therefore lacked credibility, for example; Q20 asked how the training package will improve the professionals' practise, but at the point of filling out the form they had not fully had a chance to do this yet. Furthermore, changes to professionals' knowledge and techniques is self-reported, such as Q16 that asks whether they feel they have developed skills of working effectively with service users with a personality disorder diagnosis. A more practical element may better evaluate this and remove self-report bias.

When considering the aim of the KUF awareness programme and the principles the modules are guided by to achieve this (see KUF, 2013), it is unclear whether the programme fully meets these, due to the weaknesses in the evaluation form described above. It is clear that the professionals felt more knowledgeable about the likely presentation of a personality disorder and the potential historical background of a service user that may be presenting with this disorder, as well as being able to empathise more with this client group. However, it is unclear whether this knowledge was able to be transferred to the clinical setting and

whether professionals were able to develop self-awareness and critical reflection skills. An understanding of organisations did not come through in the evaluation forms, other than to highlight the lack of commitment and time provided by the local trust for these sorts of training exercises. It was therefore hard to clarify what, if any, changes staff would realistically be able to make to their working practise and whether any changes to perceptions and attitudes were upheld.

How the findings will be disseminated

The findings were disseminated to the Clinical Psychologist within the Community Forensic Service and to commissioners at the end of 2013. The report has also been written up for journal publication.

Service impact achieved by the research and future plans

The evaluation indicated that several amendments needed to be made to the KUF training programme to ensure quality of facilitation and materials:

- Fixing technological difficulties with the KUF online programme
- A review of the materials with a view to potentially condensing some areas to prevent repetition, which may then provide the space to go into more depth on other topics
- Ensuring that training facilitators have the appropriate skills to manage large groups, rein in discussions and ensure everyone's emotional safety
- Greater commitment from the local trust to release staff for training, as well as increased organisational containment, supervision and support. Accessibility to the online training materials after completion of the course, or course hand-outs, may also help to alleviate this pressure somewhat.
- A review of the evaluation form:
    - Address the imbalance of questions focusing on online / face-to-face training
    - Allow for a qualitative section for the evaluation of the face-to-face training
    - Changes to the formatting of the form to allow fuller analysis of implementation of skills and/or an increase in knowledge and techniques that can be transferred into practise, for example; a three-month follow-up questionnaire, a practical element at the end of the training programme (e.g., individual discussion of a vignette or an

exercise challenging personality disorder myths), or pre- and post-measures.

References

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24,* 105-112.

Jorm, A. F. (2000). Mental health literacy: Public knowledge and beliefs about mental disorders. *British Journal of Psychiatry, 177;* 396-401.

KUF (2013). Retrieved August 21, 2013, from http://www.personalitydisorder.org.uk/training/kuf/

Markham, D. (2003). Attitudes towards patients with a diagnosis of 'borderline personality disorder': Social rejection and dangerousness. *Journal of Mental Health, 12(6);* 595-612.

Newton-Howes, G., Weaver, T., & Tyrer, P. (2008). Attitudes of staff towards patients with personality disorder in community mental health teams. *Australian and New Zealand Journal of Psychiatry, 42(7);* 572-577.

Trainee's Signature: _____ **Date:** _____

Supervisor's Signature: _____ **Date:** _____